Technical Report 1205

Concurrent Validation of Experimental Army Enlisted Personnel Selection and Classification Measures

Deirdre J. Knapp (Editor)
Human Resources Research Organization

Trueman R. Tremble (Editor)
U.S. Army Research Institute

August 2007

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the Department of the Army Deputy Chief of Staff, G1

Authorized and approved for distribution:

PAUL A. GADE
Acting Technical Director

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical reviews by

Richard Hoffman, U.S. Army Research Institute
Kimberly Owens, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
August 2007	Interim Report	December 1, 2003 – April 30, 2007

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Concurrent Validation of Experimental Army Enlisted Personnel Selection and Classification Measures	DASW01-03-D-0015, DO 0006
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Knapp, Deirdre J. (Editor) (Human Resources Research Organization) and Trueman R. Tremble (Editor) (U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 257
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 400 Alexandria, Virginia 22314	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1205

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Trueman R. Tremble

14. ABSTRACT *(Maximum 200 words)*: *New Predictors for Selecting and Assigning Future Force Soldiers (Select21)* is concerned with Soldier accession and job classification. The Select21 goal is to ensure the Army acquires Soldiers with the knowledge, skills, and attributes (KSAs) needed for performing well and fitting well in a transformed Army. The objectives of the project are to (a) identify Future Force job demands and the pre-enlistment KSAs required to meet them, (b) develop measures of job performance and critical KSAs, and (c) validate the experimental predictor (KSA) measures in a concurrent criterion-related validation. This report documents the method and results of the criterion-related validation.

The predictor set includes measures of cognitive ability, temperament, psychomotor skills, values, expectations, and experience. Performance criteria include rating scales completed by supervisors and peers, technical knowledge tests, a situational judgment test, and indicators of person-environment fit (e.g., job satisfaction). Versions of these measures suitable for all first-term Soldiers regardless of specialty were administered to 812 Soldiers. Analyses indicated that scores from the Armed Services Vocational Aptitude Battery (ASVAB) predict both current and future performance (as assessed by future-oriented rating scales) and that the experimental predictors provided incremental validity, particularly in regard to attitudinal criteria.

15. SUBJECT TERMS

Behavioral and social science	Personnel	Criterion-related validation
Selection and classification	Manpower	

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	291	Ellen Kinzer Technical Publications Specialist (703) 602-8047

Standard Form 298

Technical Report 1205

Concurrent Validation of Experimental Army Enlisted Personnel Selection and Classification Measures

Deirdre J. Knapp (Ed.)
Human Resources Research Organization

Trueman R. Tremble (Ed.)
U.S. Army Research Institute

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

August 2007

Army Project Number
622785A790

Personnel, Performance
and Training Technology

Acknowledgements

There are a large number of individuals not listed as authors who have contributed significantly to the work described in this report. In particular, we would like to acknowledge the contributions of Dr. John Campbell, our principal scientist, and the members of our Scientific Review Panel: Drs. Wally Borman, Fred Oswald, Bruce Orvis, Ken Pearlman, and Ben Schneider. We also want to thank Dr. Michael G. Rumsey of the U.S. Army Research Institute for his continued support of the Select21 and related research programs.

We are also indebted to the military and civilian personnel who supported our data collection efforts, particularly those Soldiers and noncommissioned officers (NCOs) who completed the experimental measures in the concurrent validation.

CONCURRENT VALIDATION OF EXPERIMENTAL ARMY ENLISTED PERSONNEL SELECTION AND CLASSIFICATION MEASURES

EXECUTIVE SUMMARY

Research Requirement:

The Select21 project was undertaken to help the U.S. Army ensure that it acquires Soldiers with the knowledges, skills, and attributes (KSAs) needed for performing the types of tasks envisioned in a transformed Army. This transformation will involve development and fielding of Future Combat Systems (FCSs) to achieve full spectrum dominance through a force that is responsive, deployable, agile, versatile, lethal, and fully survivable and sustainable under all anticipated combat conditions (U.S. Army, 2001, 2002). However, Army leadership recognizes first and foremost the importance of its people – Soldiers – to the effectiveness of transformation. In this context, the ultimate objectives of the project were to (a) develop and validate measures of critical KSAs needed for successful execution of Future Force missions, and (b) propose use of these measures as a foundation for an entry-level selection and classification system adapted to the demands of the 21st century. Earlier in the Select21 project, we conducted a future-oriented job analysis (Sager, Russell, Campbell, & Ford, 2005) to support the development of criterion measures and experimental selection and classification predictor measures (Knapp, Sager, & Tremble, 2005). The present report documents the concurrent validation effort.

Procedure:

The criterion measures and experimental predictors were administered to 812 first-term enlisted Soldiers at three locations. The criterion measures included (a) job knowledge tests, (b) a criterion situational judgment test (CSJT), (c) performance ratings (covering current performance and anticipated performance under explicitly defined future conditions) collected from supervisors and peers, and (d) surveys of current job attitudes (the Army Life Survey; ALS) and expected attitudes under defined future conditions. All Soldiers completed versions of these measures suitable for first-term Soldiers regardless of military occupational specialty (MOS). We administered job specific criterion measures to Infantrymen (11B) and Signal Support Systems Specialists (25U), but the 25U sample was too small to support planned classification efficiency analyses. Therefore, data analysis work focused primarily on the extent to which each of the experimental measures were related to Army-wide performance. These analyses included estimation of incremental validity beyond the predictive power of scores from the Armed Services Vocational Aptitude Battery (ASVAB).

The experimental predictors administered in the concurrent validation included (a) two temperament measures (Rational Biodata Inventory, RBI and Work Suitability Inventory, WSI), (b) a predictor situational judgment test (PSJT), and (c) two psychomotor tests (Target Shoot and Target Tracking). There were also two measures based on person-environment fit models, the Work Values Inventory (WVI) and the Work Preferences Survey (WPS). The WVI measures preferences for various work-related reinforcers (e.g., opportunity to learn new things), whereas the WPS measures interest in various activities. Some measures developed in Select21 and described

in Knapp et al. (2005) were not included in the concurrent validation because they were not suitable for administration to experienced Soldiers (e.g., the Pre-Service Expectations Survey).

Findings:

Overall, the results of the predictor cross-instrument analyses suggest little appreciable overlap among the predictors. Although some of the measures have scales that assess similar constructs, and the correlations between these measures were significant and moderate in strength (supporting evidence for convergent validity), the magnitude of the correlations was not so high as to suggest substantial measurement redundancy. In further support of the measures' convergent and discriminant validity, correlations among scales from different instruments that purported to measure similar constructs were generally stronger than correlations with scales that were designed to measure different constructs.

Our intent was to develop predictors that supplement the ASVAB for the prediction of performance and attitudinal criteria. We constructed five composite performance scores (based on a confirmatory factor analysis modeling exercise) and five attitudinal scores to use in the validation analyses. The five performance criteria were (a) General Technical Proficiency, (b) Achievement and Effort, (c) Physical Fitness, (d) Teamwork, and (e) Future Expected Performance. The five attitudinal scores were (a) Satisfaction with the Army, (b) Perceived Army fit, (c) attrition cognitions, (d) career intentions, and (e) Future Army Affect.

Consistent with prior research, scores on the ASVAB continued to be good predictors of can-do performance criteria (e.g., General Technical Proficiency) and to have less validity for predicting will-do (e.g., Physical Fitness, Teamwork) and attitudinal criteria. ASVAB scores yielded significant correlations with future expected performance scores; this is a new finding, and one that bears emphasis. ASVAB scores yielded small but significant negative correlations with attrition cognitions. Soldiers with higher cognitive ability were less likely to think about breaking their enlistment contract.

On the other hand, many of the Select21 predictors showed notable levels of incremental validity over the ASVAB when predicting Achievement and Effort, Physical Fitness, and Teamwork performance. Such findings reinforce the notion that when judging the efficacy of predictors for incrementing the validity of the ASVAB, it is important to account for the multi-dimensional nature of the criterion space. Substantial levels of incremental validity were found for the RBI, WVI, and WPS for predicting the attitudinal criteria, with somewhat lower levels of validity for the WSI and PSJT. While findings for the RBI were quite strong for the attitudinal criteria, such results appeared to partially reflect criterion-related contamination stemming from the inclusion of the RBI Army Identification scale in the RBI predictor composite. Nevertheless, even with the Army Identification scale removed, the RBI still exhibited notable levels of incremental validity for predicting the attitudinal criteria.

We performed subgroup analyses using type of MOS as the subgrouping variable to get an idea of the potential for the experimental predictors to improve classification efficiency. Soldiers were sorted into four MOS clusters for these analyses, which did suggest that some of the predictors have potential utility for classification. Six predictor measure scales showed differences in validity estimates across clusters for three or more criterion composites: (a) RBI

Fitness Motivation, (b) WSI Attention to Detail, (c) WPS Creativity, (d) WPS Physical, (e) RBI Army Identification, and (f) Target Tracking. Other predictors showed more targeted results focused on specific cluster comparisons or criteria.

Utilization and Dissemination of Findings:

Many of the new Select21 predictors are self-report indicators in which scores may be affected by experience in the Army and response distortion (whether intentional or not) in an operational setting. Therefore, it is particularly important to evaluate them in a longitudinal validation in which the predictors are administered to Army applicants or new recruits. A follow-on 5-year research program known as "Army Class" has been initiated to collect such data. Moreover, Army Class is designed to gather more MOS-specific data from Soldiers in the 11B and 25U MOS (which can then be combined with the MOS-specific collected from Soldiers in these MOS in Select21), as well as MOS-specific data from Soldiers in a broader sampling of MOS. This will allow a more definitive assessment of the classification potential of the experimental predictors. Army Class includes a concurrent validation as well as a longitudinal validation, so it will significantly move forward the foundation provided by Select21 for implementation of new enlistment tests.

CONCURRENT VALIDATION OF EXPERIMENTAL ARMY ENLISTED PERSONNEL SELECTION AND CLASSIFICATION MEASURES

CONTENTS

CONTENTS (Continued)

Page

Page

Page

Page

List of Tables

Page

Page

Page

Page

List of Figures

CONCURRENT VALIDATION OF EXPERIMENTAL ARMY ENLISTED PERSONNEL SELECTION AND CLASSIFICATION TESTS

PART 1: BACKGROUND

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp
HumRRO

Overview of the Select21 Project

The U.S. Army is undertaking fundamental changes to transform into the Future Force. The 4-year Select21 project concerned future entry-level Soldier selection, with the goal of ensuring that the Army selects and classifies Soldiers with the knowledge, skills, and attributes (KSAs) needed for performing successfully in a transformed Army. The ultimate objectives of the project were to (a) develop and validate measures of critical attributes needed for successful execution of Future Force missions, and (b) propose use of the measures as a foundation for an entry-level selection and classification system adapted to the demands of the 21st century. The Select21 project focused on the period of transformation to the Future Force—a transition envisioned to take on the order of 30 years to complete. The time frame of interest extends to approximately 2025.

The major elements of the approach used in this project were (a) future-oriented job analysis, (b) development of predictor measures suitable for predicting performance in the future Army, (c) development of criterion measures consistent with anticipated future Army requirements, and (d) a concurrent criterion-related validation effort. The future-oriented job analysis (Sager, Russell, Campbell, & Ford, 2005) provided the foundation for the development of new tests that could be used for recruit selection or Military Occupational Specialty (MOS) assignment/ classification (i.e., predictors) and the development of job performance measures that serve as criteria for evaluating the predictors. Development of the Select21 predictor and criterion measures was documented in Knapp, Sager, and Tremble (2005). The purpose of the present report is to describe the final stage of the project—the concurrent validation procedure and results.

The Select21 research program was sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) with contract support from the Human Resources Research Organization (HumRRO). The remainder of this chapter summarizes the overall Select21 research approach, including the (a) identification of job clusters and job sampling, (b) job analysis findings, (c) criterion measures, (d) predictor measures, and (e) the concurrent validation plan. The chapter concludes with an overview of the rest of the report.

Job Clusters and Sampling

The original Select21 research plan (May, 2002) called for the identification of clusters of future Army jobs. The clusters would provide a basis for determining whether any of the experimental predictor measures had potential for improving classification decisions without relying too heavily on the Army's current job structures (i.e., MOS and associated MOS

categorizations such as Career Management Fields [CMF]). Sixteen future entry-level Army job clusters were identified (Sager et al., 2005). We selected two clusters for closer examination in the validation research: Close Combat and Surveillance, Intelligence, and Communication (SINC). The primary reasons for selecting these two clusters were that they were both considered very important to the Future Force while also being maximally distinct from each other, thus maximizing the opportunity to evaluate the classification potential of the predictor measures.

The plan was for the concurrent validation to include multiple research samples— an Army-wide sample (with Soldiers drawn from all MOS without regard to cluster membership) and several MOS-specific samples drawn from two job clusters. Therefore, we collected job analysis information for Army-wide requirements (applicable to all MOS) and for six individual MOS representing the two target job clusters (see Table 1.1). The Army-wide job analysis information was intended to support design and development of predictors and criteria suitable for selection-based research (i.e., selecting new recruits) and the MOS/job cluster analysis information was intended to support classification-based research (i.e., assigning new recruits to Army jobs). Although we collected some limited cluster-level information, the job analysis information required to support most criterion work required us to focus on the MOS level.

Table 1.1. Select21 Target Job Clusters and MOS

Close Combat		
	11B	Infantryman
	19D	Cavalry Scout
	19K	M1 Armor Crewman
Surveillance, Intelligence, and Communications (SINC)		
	25U	Signal Support Systems Specialist (formerly 31U)
	25B	Information Systems Operator/Analyst (formerly 74B)
	96B	Intelligence Analyst

Job Analysis Findings

The Select21 job analysis work characterized future entry-level Army enlisted job requirements in several complementary ways. Job requirements were defined in terms of the following:

- Performance Requirements
  - Performance dimensions (Army-wide)
  - Common tasks (Army-wide)
  - Job tasks/task categories (for each target MOS)
  - Anticipated future conditions (Army-wide and for each target job cluster)
- Pre-enlistment KSAs (Army-wide, prioritized by MOS)

The procedure for conducting the future-oriented job analysis is described in detail in Sager et al. (2005). Our interest was in job requirements for fully trained Soldiers serving their first

enlistment term. Given the Army's training system, we defined an entry-level Soldier as one in his or her first enlistment term, with 18-36 months time-in-service.

Performance Requirements

We used information available from existing resources as a starting point for defining performance requirements. These sources included Army occupational analysis findings, training manuals, prior research (e.g., NCO21 and Project A; Campbell & Knapp, 2001; Ford, Campbell, Campbell, Knapp, & Walker, 2000), and information from the Future Force literature. Project staff developed draft materials which were then subjected to an iterative review and revision process involving subject matter experts (SMEs) familiar with the Future Force vision and/or their own MOS. This process involved a series of workshops which resulted in detailed descriptions of Army-wide and MOS requirements for the six target MOS.

Specifically, the job analysis process yielded a list of 19 Army-wide performance dimensions and 59 Army-wide common tasks (Sager et al., 2005). It also produced task lists (organized into categories) for the six MOS representing the Close Combat and SINC clusters. The performance dimensions and job tasks are provided in Sager et al. (2005).

Unlike typical job analyses that focus on current job requirements, it was important to capture information about the context of performance in the future Army. That is, the conditions in which Soldiers will be performing needed to be made explicit to help support development of criterion measures that, inasmuch as possible, reflect future-oriented performance. Table 1.2 lists the anticipated future conditions for all entry-level Soldiers in the Future Force. MOS/cluster-specific future conditions were also identified and are provided in Sager et al. (2005).

Table 1.2. Army-Wide Anticipated Future Conditions

Learning Environment: Greater requirement for continuous learning and the need to independently maintain/increase proficiency on assigned tasks.
Disciplined Initiative: Less reliance on supervisors and/or peers to perform assigned tasks.
Communication Method and Frequency: Greater need to function based on digitized instead of face-to-face communication; greater understanding of the common operational picture and increased situational awareness.
Individual Pace and Intensity: Greater need for mental and physical stamina and greater awareness of one's own mental and physiological status; greater task variety.
Self-Management: Greater emphasis on ensuring that Soldiers balance and manage their personal matters and well-being.
Survivability: Improved protective systems, transportation, communication, and medical care will result in an incremental improvement in personal safety.

Pre-Enlistment KSAs

As with the performance requirements, the job analysis team reviewed multiple available sources to generate a list of potentially applicable pre-enlistment KSAs. As described by Sager et al. (2005), these sources included the Basic Combat Training list, Project A KSAs, NCO21 KSAs, as

well as the relevant psychological research literature. This activity resulted in a list of 48 KSAs relevant to performance of first-term Soldiers in the Future Force. The list was reviewed by Army SMEs and the Select21 Scientific Review Panel (a group of preeminent researchers who periodically reviewed the Select21 research activities). SMEs prioritized the pre-enlistment KSAs by importance for all Soldiers Army-wide and for Soldiers in each target MOS.

Criterion Measurement Plan

Our goal was to develop criterion measures that, taken together, would provide reasonably comprehensive coverage of the criterion space in terms of content and scores that reflect all performance determinants (i.e., declarative knowledge, procedural knowledge and skills, and motivation) (Campbell, McCloy, Oppler, & Sager, 1993). A guide for such coverage was the performance model developed in Project A (Campbell & Knapp, 2001). In Project A, first-term Soldier performance was characterized by a model with five factors: Core Technical Proficiency, General Soldiering Proficiency, Effort and Leadership, Maintaining Personal Discipline, and Physical Fitness and Military Bearing. We also sought to address issues of Soldier retention by including criterion measures reflecting a person's fit within the work/organizational environment. These person-environment (P-E) fit measures include items related to such constructs as job satisfaction and organizational commitment.

A particularly challenging goal of the Select21 criterion measures was for them to reflect how well Soldiers would perform in the Future Force. Obviously, this is something that must be approximated as closely as possible rather than being a fully achievable goal. We used the following strategies to examine future performance and organizational fit:

- Base the content of criterion tests on future-oriented job analysis results.
- Provide respondents (raters and Soldiers) with a basis for making predictions about the future.

To meet our goals, the Select21 criterion measures thus included the following:

- Performance rating scales covering both current and expected future performance (completed by supervisors and peers)
- Job knowledge tests
- Archival/self-report information (e.g., military training, disciplinary actions, attrition)
- A criterion situational judgment test (CSJT)
- A self-report measure of job satisfaction and organizational fit (Army Life Survey)

Figure 1.1 depicts how these criterion measures correspond to the 19 Army-wide performance dimensions identified in the job analysis.

Performance Ratings

Although subjective ratings tend to exhibit a number of problems when used as criterion measures, they can comprehensively tap important dimensions of performance and can also provide perhaps the best indicator of typical (versus maximal) performance. In Select21, we

developed rating scales and data collection procedures intended to maximize the information obtained using this measurement method (i.e., efficient and comprehensive measurement of the performance space) while minimizing the disadvantages (e.g., reliance on human raters who are prone to rating error such as halo and leniency bias).

Army-Wide Performance Dimensions	Rating Scales[a]	Job Knowledge Tests[b]	CSJT	Archival/ Self-report
Performs Common Tasks	X	X		
Solves Problems/Makes Decisions	X			
Exhibits Safety Consciousness	X	(X)[c]		
Adapts to Changing Situations	X		X	
Communicates in Writing	X			
Communicates Orally	X			
Uses Computers	X			(X)[c]
Manages Information	X			
Exhibits Cultural Tolerance	X			
Exhibits Effort and Initiative on the Job	X			(X)[c]
Follows Instructions and Rules	X	(X)[c]		X
Exhibits Integrity and Discipline on the Job	X			(X)[c]
Demonstrates Physical Fitness	X			X
Demonstrates Military Presence	X			
Relates to and Supports Peers	X		X	
Exhibits a Selfless Service Orientation	X			(X)[c]
Exhibits Self-Management	X		X	
Exhibits Self-Directed Learning	X		X	
Demonstrates Teamwork	X		X	

Note. The Army Life Survey is not listed because it was not designed to cover these performance dimensions.
[a]MOS-specific rating scales covered MOS-specific task categories; the Future Expected Performance Rating Scales covered the anticipated future conditions.
[b]The job knowledge tests covered both Army-wide (common) and MOS-specific tasks.
[c]Parentheses indicate indirect assessment of the performance dimension.

Figure 1.1. Select21 criterion measures by performance dimensions matrix.

We developed two types of rating scales designed to be completed by both supervisors and peers. One set of scales (the Current Observed Performance Rating Scales) requires raters to consider current observed performance whereas the other set of scales (Future Expected Performance Rating Scales) requires raters to estimate performance under conditions expected to characterize the future Army. The rating scale format, training, and rating procedures were designed to (a) minimize rater errors, (b) focus the raters on the rating scale dimension definitions and anchors, (c) help raters differentiate between performance in the current Army and performance in the future Army, and (d) facilitate the collection of complete ratings data on all target Soldiers. Our goal was to collect one supervisor rating and three peer ratings per Soldier.

Job Knowledge Tests

Job knowledge tests were selected as the primary means for measuring task proficiency. Hands-on tests, which would have provided a more direct measure of task proficiency, were not used because of the resources required to administer them. Although job knowledge tests are

lower fidelity assessments compared to hands-on tests, they do offer the advantage of relatively comprehensive task coverage. Moreover, Select21 test developers used a variety of item formats (e.g., multiple-choice, drag and drop, ranking, matching) and graphics to enhance the realism of these computer-administered tests as well as minimize reading requirements. Project staff drafted tests (one Army-wide and one for each target MOS) using test blueprints based on the Select21 job analysis results and SME input. Because these tests cover detailed knowledge of how to perform current job tasks (and comparable information cannot be known for future job tasks), they are not future performance measures, *per se*. The test blueprints are, however, based on findings from the future-oriented job analysis. Furthermore, although the demands for acquiring knowledge might increase in the future (as indicated by the Select21 future-oriented job analysis), there is little reason to believe that the ability to acquire declarative knowledge in the future will be predicted by different KSAs than the ability to acquire such knowledge today.

Criterion Situational Judgment Test (CSJT)

In prior research, several of the Army-wide performance dimensions have been successfully embedded in situational judgment tests (e.g., Campbell & Knapp, 2001; Knapp, Burnfield et al., 2002). The Select21 Criterion Situational Judgment Test (CSJT) presents problem scenarios common to Soldiers reaching the end of their first terms of enlistment, along with several possible response options. Test scores are computed by comparing Soldier responses with "expert" responses (judgments) made by a sample of senior noncommissioned officers (NCOs). As with the job knowledge tests, the dimensions covered by the CSJT are based on the Select21 future-oriented job analysis.

Archival/Self-Report Information

Variations of the Personnel File Form have been used in several ARI research projects since it was originally developed for Project A (Campbell & Knapp, 2001). The form draws much of its content from the Army's enlisted personnel "Promotion Point Worksheet." Obtaining the information via self-report is quick, accurate, and efficient (Riegelhaupt, Harris, & Sadacca, 1987) and allows collection of additional information that would not otherwise be readily accessible (e.g., recent disciplinary actions). By its nature, the archival/self-report information reflects performance under current Army conditions.

Although the Select21 project relied on a concurrent research design that did not allow collection of archival attrition data from the primary validation sample, considerable data were collected from new recruits in the development and field testing of the predictor measures in 2003-2004. During the timeframe of this project, then, it was possible to examine the relationship between Select21 predictors and attrition from basic training, advanced training, and (for some research participants) operational units. This work was conducted somewhat independently from the primary research effort, so it is documented more thoroughly elsewhere (e.g., Putka & Le, 2005).

Army Life Survey (ALS)

One goal of Select21 was to expand the criterion space to include attrition (separation from the Army prior to completion of the first enlistment term) and retention (remaining in the Army beyond the initial enlistment term). As discussed previously, the concurrent validation

research design did not allow for examination of these behaviors in our primary research sample. Instead, we developed person-environment fit indicators that are theoretical precursors to turnover behaviors. The Army Life Survey (ALS) was developed to measure job satisfaction, organizational commitment, perceived stress, perceived fit, turnover intentions, and perceived importance of core Army values. The Future Army Life Survey (FALS) is a shorter instrument that describes various aspects of the Army of the future and asks Soldiers to indicate how these aspects would affect their feelings toward the Army.

Predictor Measurement Plan

A fundamental goal of the Select21 project was to determine the possibility of developing selection and classification measures that (a) predict the performance of entry-level Soldiers in the Future Force and (b) add incremental validity over the current system as embodied by the Armed Services Aptitude Battery (ASVAB). The measures we developed were designed to cover the KSAs identified in the Select21 job analysis.

The Select21 measures for predicting future performance included the following:

- Armed Services Vocational Aptitude Battery (ASVAB)
- Temperament measures
  - Rational Biodata Inventory (RBI)
  - Work Suitability Inventory (WSI)
- Psychomotor measures
  - Target Shoot
  - Target Tracking
- Predictor situational judgment test (PSJT)
- Record of Pre-Enlistment Training and Experience (REPETE)

Figure 1.2 shows the coverage these instruments provide of the Select21 pre-enlistment KSAs. Note that not all KSAs are covered. In particular, the measures did not cover KSAs related to physical abilities (e.g., static strength, dynamic flexibility) that represent medical or physical fitness domains outside the scope of ARI's mission. Note also that, as with the criterion measures, each instrument was not designed to produce scores specific to each KSA. Rather, the content of the instruments was designed to reflect the subset of KSAs noted in the figure. Finally, Figure 1.2 does not include the P-E fit instruments because they were not designed to cover KSAs, *per se*. The P-E fit predictor measures were as follows:

- Work Values Inventory (WVI)
- Work Preferences Survey (WPS)
- Career Exploration Program Interest Inventory (CEP-II)[1]
- Army Beliefs Survey (ABS)
- Pre-Service Expectations Survey (PSES)
- Army Work Knowledge Survey (AWKS)

[1] The CEP-II was developed by the Defense Manpower Data Center (DMDC) and was used primarily as a marker measure for the WPS.

KSA	ASVAB	RBI	WSI	PSJT	Psychomotor	REPETE
Oral Communication Skill						
Oral and Nonverbal Comprehension						
Written Communication Skill						
Reading Skill/Comprehension	✓					
Basic Math Facility	✓					
General Cognitive Aptitude	✓					
Spatial Relations Aptitude	✓					
Vigilance						
Working Memory						
Pattern Recognition						
Selective Attention						
Perceptual Speed and Accuracy						
Team Orientation			✓	✓		
Agreeableness		✓	✓	✓		
Cultural Tolerance		✓	✓			
Social Perceptiveness			✓	✓		
Achievement Motivation		✓	✓	✓		
Self-Reliance		✓	✓			
Affiliation		✓	✓			
Potency		✓	✓			
Dependability		✓	✓	✓		
Locus of Control		✓				
Intellectance		✓	✓			
Emotional Stability		✓	✓			
Static Strength						
Explosive Strength						
Dynamic Strength						
Trunk Strength						
Stamina						
Extent Flexibility						
Dynamic Flexibility						
Gross Body Coordination						
Gross Body Equilibrium						
Visual Ability						
Auditory Ability						
Multilimb Coordination					✓	
Rate Control					✓	
Control Precision					✓	
Manual Dexterity						
Arm-Hand Steadiness						
Wrist, Finger Speed						
Hand-Eye Coordination						
Basic Computer Skill						✓
Basic Electronics Knowledge	✓					✓
Basic Mechanical Knowledge	✓					✓
Self-Management Skill				✓		
Self-Directed Learning and Development Skill				✓		
Sound Judgment				✓		

Note. The P-E fit measures are not included because they are not designed to assess KSAs.

Figure 1.2. Select21 predictor measures by KSA matrix.

Baseline Predictors

The current selection and classification system relies largely on the ASVAB. Thus, the ASVAB served as the baseline against which the Select21 experimental predictors were compared. The ASVAB contains one experimental subtest—Assembling Objects (AO)—and nine operational subtests. Applicants must meet a minimum score on the Armed Forces Qualification Test (AFQT) that is a composite of four ASVAB subtests to enter the Army. For MOS assignment, the applicants' ASVAB scores must meet minimum qualifying scores set for each MOS. Another baseline predictor used in Select21 was educational status (i.e., high school diploma status), which is used by the Army to predict attrition. ASVAB scores and pre-enlistment educational tier were retrieved from Soldier personnel records for use in the Select21 research.

Temperament Measures

Prior research has shown that the ASVAB is a psychometrically strong measure of cognitive aptitude and an effective predictor of job performance in general and task proficiency in particular. Thus, the experimental predictors developed for Select21 emphasized non-cognitive characteristics likely to predict the more motivational aspects of performance and turnover (i.e., attrition and reenlistment behavior). Several of the temperament-based measures described below used different approaches to try to tackle the problem of response distortion (i.e., faking) that has long daunted personnel psychologists.

Rational Biodata Inventory (RBI)

The RBI is an instrument that, in various forms, has been used in prior Army research and operational applications (e.g., for selection into Special Forces) for several years. As its name suggests, the RBI is a self-report measure that uses Likert-style response options. It yields scores on several substantive areas (e.g., Achievement Motivation, Hostility to Authority), and also includes a response distortion scale. The idea behind the response distortion scale is that scores on this scale can be used to identify individuals whose scores on the other RBI scales are suspect, and such scores can then be adjusted accordingly. Moreover, over the course of instrument development, the response distortion scores were used to eliminate items that appeared particularly subject to distortion.

Work Suitability Inventory (WSI)

The WSI asks respondents to rank order statements that describe different work styles. Each work style statement corresponds to a temperament construct. Using items that reflect work preferences rather than temperament *per se* is one strategy the WSI uses to combat response distortion. Because it is a ranking task (which also minimizes response distortion), the one-item dimension-level scores are fully ipsative. In other words, the dimension-level scores constrain each other (e.g., if you are high on one dimension you must be lower on another) making it difficult to compare scores across individuals. The ipsativity problem is mitigated, however, by the construction of one or more empirically-derived composite scores (using subsets of the dimension-level scores) geared to the prediction of a given criterion. The idea is that the Army could construct multiple composite scores, each using a different array of dimensions that are

geared to the prediction of various pre- and post-enlistment criteria (e.g., attrition, performance as a Drill Sergeant, performance as a recruiter). These composite scores would be potentially useful as a basis for personnel decisions.

Predictor Situational Judgment Test (PSJT)

In addition to being used for performance measurement, the situational judgment test method has often been used to develop effective predictor measures (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). Given its past effectiveness, we developed an experimental predictor based on this method. The instrument consists of civilian problem scenarios that parallel situations experienced by Soldiers during their first few months in the Army. Project researchers experimented with several ways to score the PSJT, including one method that would yield temperament-like (i.e., trait) scores. If such a scoring strategy were successful, the PSJT could provide another strategy for assessing temperament that deals with response distortion in a way that is distinct from the RBI and WSI.

Psychomotor Tests

Prior research has shown that psychomotor tests can be useful for classifying Army applicants into MOS (Campbell & Knapp, 2001), but previously the technology for large-scale psychomotor testing was limited. Given advances in this technology, Select21 researchers adapted two psychomotor tests originally developed in Project A (Campbell & Knapp, 2001). The two tests are Target Shoot and Target Tracking.

Record of Pre-Enlistment Training and Experience (REPETE)

Historically, the Army has assumed the burden of training all required entry-level job skills for its enlisted personnel. Recognizing prior training and/or experience could benefit the Army by reducing training requirements (or at least helping to ensure success in training) and could also benefit applicants by enhancing their enlistment options (in terms of job choices and/or enlistment bonuses). Such a tool could also be particularly helpful in accessioning new Soldiers (e.g., reserve component Soldiers, personnel moving from other services) who more likely have pertinent job skills prior to entry.

Based on this hypothesis, the Select21 project developed a self-report experimental predictor measure to determine what types of training and experience entry-level Soldiers bring with them to the Army. To develop this measure, project staff reviewed all the Select21 KSAs and constructed questions that query respondents about related training, certifications, and experience. Particular attention was given to computer-related skills. The field-tested version of the REPETE helped demonstrate the potential value of this type of measure (Russell, Le, & Knapp, 2005). However, it was not included in the concurrent validation because we believed it would be too difficult for Soldiers who had been in the Army for 18-36 months to report detailed pre-enlistment training and experience accurately.

P-E Fit Predictors

We developed several experimental predictors based on the concept of person-environment fit. The *Work Values Inventory (WVI)* uses a ranking exercise to determine what characteristics of work situations are particularly important to an individual (e.g., the opportunity to work with people, having clearly defined work requirements). The *Work Preferences Survey (WPS)* assesses an individual's work-related interests. Unlike most interest inventories, the WPS was designed for selection and classification rather than to support career counseling. We also administered the Career Exploration Program Interest Inventory (CEP-II), a measure used by the Department of Defense to support career counseling for high school students.

A set of three P-E fit predictors was developed based on the idea that applicants who have realistic expectations about the Army prior to enlistment will have a greater chance of being satisfied with the Army and staying in the Army at least through their first enlistment term. We created these measures by taking the content from the WSI (i.e., work styles), WVI (i.e., work characteristics), and WPS (i.e., interests) and asking respondents to indicate the degree to which each is characteristic of their MOS. The instruments were scored by comparing the respondent's answers to the average of comparable responses provided by non-commissioned officers (NCOs) in the MOS. These measures (the Army Beliefs Survey, the Pre-Service Expectations Survey, and the Army Work Knowledge Survey) showed promise during field testing (Van Iddekinge, Putka, & Sager, 2005), but were not suitable for administration in a concurrent validation. As with the REPETE, the concern was that the retrospective responses of experienced Soldiers would not accurately reflect the responses they would have given at service entry.

Concurrent Validation

The experimental predictors and the criterion measures were administered to Soldiers during 2005 and very early in 2006. Our goal was to collect data from Soldiers in their first enlistment term with 18-36 months time in service, with the idea that this would approximate late first-term performance across Soldiers with different enlistment terms and varying lengths of training prior to being sent to their units. As discussed in the technical report documenting development of the predictor and criterion measures (Knapp et al., 2005), it was quite difficult to obtain adequate numbers of Soldiers in the six target MOS. Therefore, the research plan was modified to represent the two target MOS clusters with one MOS each instead of three. Specifically, we targeted three samples of Soldiers in the concurrent validation—(a) an Army-wide (mixed MOS) sample, (b) an infantry (11B) sample representing the Close Combat MOS cluster, and (c) a signal support systems specialist (25U) sample representing the Surveillance, Intelligence, and Communications (SINC) cluster.

As described further in Chapter 2, we collected data on a total of 812 Soldiers. This includes 539 Soldiers in the Army-wide sample. It also includes 216 Soldiers in the 11B MOS sample for whom we collected MOS-specific criterion data. Despite our best efforts, however, we collected concurrent validation data from just 57 Soldiers in the 25U MOS sample. The 25U sample size was insufficient for estimating classification gains, so this report does not include analyses related to the prediction of MOS-specific criteria. Instead, we examined the potential for the experimental predictor measures to support classification decisions by examining

differential prediction of the Army-wide criteria for subgroups of MOS across the entire sample of 812 Soldiers (see Chapter 14). Moreover, current plans are to combine the 11B and 25U data with comparable concurrent validation data being collected in 2006 as part of a follow-on project (HumRRO, 2006). In this way, we hope to achieve sufficient sample sizes for these and other MOS to support estimates of classification potential for the experimental predictors using MOS-specific criterion scores.

Overview of Report

This report is organized into five major sections. Part 1 (Background) includes this chapter and Chapter 2. Chapter 2 describes the concurrent validation data collection and resulting research sample. Part 2 (Validation Criteria), which includes Chapters 3 through 5, describes the Select21 criterion measures starting with the attitude-related scores, followed by the performance-related scores, and then examines relations among the criterion scores used in the validation analyses described in the remainder of the report. Part 3 (Individual Predictors and Bivariate Validity Results) includes Chapter 6 through 12, which describe the validation results associated with each predictor measure used in the concurrent validation. Chapter 6 reports results for the ASVAB, and subsequent chapters in this section provide incremental validity estimates beyond that provided by the AFQT composite. Chapter 6 also provides a detailed description of the analytic approach used throughout the report to examine zero-order validity, incremental validity, subgroup differences, and differential prediction. Part 4 (Predictor Intercorrelations and Multivariate Validation Results) includes Chapters 13 and 14. Chapter 13 summarizes relations among the predictor scores and examines the incremental validity of the full battery of predictors over AFQT and other selected ASVAB scores. Chapter 14 examines the validity of the various experimental predictors when computed on subgroups of the total sample defined by their MOS representation. These MOS clusters correspond to a subset of the 16 future MOS clusters identified in the Select21 job analysis work (Sager et al., 2005). Part 5 of the report concludes with a single chapter (Chapter 15) which summarizes the results of the Select21 concurrent validation research, provides commentary about the research, and offers suggestions for future research.

A companion report will focus on attrition and its prediction by the Select21 measures. Results of both the present report and the attrition report will provide the empirical bases for a final report on recommendations regarding the use of the Select21 experimental predictors.

CHAPTER 2: CONCURRENT VALIDATION DATA COLLECTION AND DATABASE DEVELOPMENT

Christopher E. Sager and Ani S. DiFazio[2]
HumRRO

Introduction

This chapter describes the Select21 concurrent validation data collection, construction of the analysis database, and sample sizes. Data were collected at three Army installations on four occasions from April 2005 to January 2006. Participants included 813 first-term enlisted Soldiers and 388 supervisors who provided performance ratings for 700 of these Soldiers.

Soliciting Participation

The commands at three installations provided research support for the concurrent validation. In securing support, ARI requested participation by first-term enlisted Soldiers and at least one supervisor per participating first-term Soldier. The support request defined "first-term Soldier" as a Soldier serving in his/her first term of service and as having completed between 18 and 36 months time in service (TIS). The tenures of some Soldiers who appeared for data collection sessions, however, were outside those specified by ARI's request. Because the installations were having trouble meeting the numbers of requested Soldiers, we expanded the pool of eligible participants. Specifically, we modified the rule for participation such that any Soldier satisfying one of two conditions was eligible for participation: (a) between 12 and 36 months TIS or (b) currently in his or her first enlistment (if the Soldier had more than 36 months TIS). Additionally, we accepted individuals with less than 12 months TIS if we had room on that day with the idea that we would later determine if their data should be included in the validation analyses. Inclusion of such Soldiers (11% of the total sample) does not appear to have been a problem for the validation analyses. That is, correlations between predictors and criteria partialling out TIS were not appreciably different from the comparable zero-order correlations between these variables.

In securing research support, the project also requested participation by three types of first-term Soldiers. These types correspond to the concurrent validation research plan. The plan called for Soldiers representing two specific MOS—11B and 25U. The plan also called for an Army-wide (AW) sample; that is, Soldiers distributed across MOS but not serving in the two specifically targeted MOS.

Table 2.1 shows the dates and numbers of Soldiers and supervisors participating at each site visit.[3] As can be seen in the table, one installation (Fort Hood) provided most of our participants. The table sorts the obtained sample into "waves." This report refers to the data

[2] We would like to acknowledge the following individuals who worked tirelessly during one or more of the data collection site visits. The ARI staff included Robert Killcullen, Kimberly Owens, Jennifer Solberg, and Trueman Tremble. The HumRRO staff included Roy Campbell, Daniel Furr, Patricia Keenan, Arthur Paddock, Dan Putka, Masayu Ramli, Teresa Russell, Megan Shay, Mary Warthen, Gordon Waugh, and Shelly West.

[3] The samples sizes in Table 2.1 represent the number of participants who completed Soldier or Supervisor Background information forms. Notes from the session logs indicated 813 Soldier participants; however, only 812 Soldier Background Information Forms were completed.

collected during the first two site visits (i.e., Forts Drum and Hood-I) as the Wave 1 data set. Data collected during the two later site visits are referred to as the Wave 2 data set. The combination of Waves 1 and 2 are referred to as the full data set. Some later chapters describe how preliminary predictor scoring and criterion-space modeling analyses took advantage of the earlier available Wave 1 data set. We would have preferred conducting these preliminary analyses on a random sample selected from the full data set. However, this approach was dictated by time constraints and the need to report preliminary validation results before the Wave 2 data collection and processing was complete. This was judged to be an acceptable alternative given the demographic similarities between the Wave 1 and 2 samples.

Table 2.1. Soldier Participation by Site Visit

Wave	Installation	Dates	Number of Participating Soldiers	Number of Participating Supervisors
1	Fort Drum	20 - 21 April 05	57	19
1	Fort Hood-I	11 - 22 July 05	572	240
2	Fort Hood-II	12 - 15 December 05	131	92
2	Fort Gordon	9 - 12 January 06	52	37
	Total		812	388

Note. Fort Hood-I and Fort Hood-II refer to two site visits to the same installation. The numbers for supervisors include participants who completed their ratings on site and those who mailed them back later.

On-Site Data Collection Procedures

Data collectors arrived at each site one or two days before sessions began to coordinate with the site point-of-contact (POC) and to set up the testing rooms. Set up procedures included preparing the available space and equipment for (a) Soldier paper-and-pencil sessions, (b) Soldier computerized sessions, and (c) supervisor paper-based performance rating sessions. Soldier participation lasted for a day (i.e., a paper-and-pencil session and a computerized session). Supervisors were asked to arrive with their Soldiers in the morning for a rating session that lasted about 1.5 hours; however, supervisors were accommodated at any time during the day.

Soldier Sessions

The day-long data collection period for Soldiers was divided into a computerized session and a paper-and-pencil session for Soldiers. Each of the two types of Soldier sessions was scheduled to last for 4 hours. Table 2.2 shows the instruments administered in the computer session. Table 2.3 shows the instruments administered in the paper-and-pencil session. If more than 25 Soldiers attended, they were split into two groups (i.e., one started with the paper and pencil session and the other started with the computerized session). The tables also show the instruments in their order of administration. The superscripts in Tables 2.2 and 2.3 indicate which instruments were exclusive to either the AW or the MOS-specific Soldiers.

At the beginning of their morning session, all Soldiers (a) completed a sign-in sheet, (b) filled out a Supervisor and Peer Rater Identification Sheet, (c) listened to a project briefing, and (d) completed a Soldier Background Information Form. The Background Information Form included a Privacy Act Statement and required the Soldier to provide identification and demographic information (e.g., social security number [SSN], pay grade, MOS, gender, and race).

Table 2.2 Instruments Administered in Soldier Computer Sessions

- Personnel File Form
- Army-Wide Job Knowledge Test
- Work Suitability Inventory
- Work Values Inventory
- Criterion Situational Judgment Test[a]
- MOS-Specific Job Knowledge Test[b]
- Psychomotor Tests

[a] Initially this instrument was administered only during AW sessions; later (once it was evident that there was sufficient time to do so) it was also administered during MOS-specific sessions.
[b] 11B and 25U versions of this instrument were administered only during the MOS-specific sessions.

Table 2.3 Instruments Administered in Soldier Paper-and-Pencil Sessions

- Rational Biodata Inventory
- Work Preferences Survey
- Predictor Situational Judgment Test
- Career Exploration Program Interest Inventory[a]
- Army Life Survey
- Peer Ratings
  - Army-Wide Current Observed Performance Rating Scales
  - MOS-Specific Current Observed Performance Rating Scales[b]
  - Army-Wide Future Expected Performance Rating Scales
  - MOS-Specific Future Expected Performance Rating Scales[b]
- Future Army Life Survey

[a] This instrument was administered only during AW sessions.
[b] 11B and 25U versions of this instrument were administered only during the MOS-specific sessions.

The Rater Identification Sheet required Soldiers to identify (a) two supervisors who could rate their performance, (b) up to four peers who could rate their performance, and (c) up to four peers whose performance they could rate. Soldiers needed to have worked with all nominees for at least a month and peer nominees needed to be participating in the data collection that day. Based on this information provided by the Soldiers, a custom-made ACCESS program was used to match peer raters to eligible ratees with the goal of maximizing the number of raters per ratee and ensuring that no Soldier was required to rate more than four peers. At the beginning of the peer rating process, each Soldier was given a rating card listing the names and identification numbers of peers to be rated. The Soldiers then underwent training for the current performance ratings (see Keenan, Russell, Le, Katkowski, & Knapp [2005] for a description of the rater training for the Current Observed Performance Rating Scales). The training included (a) familiarization with the performance dimensions and their anchored rating scales, (b) description of common rating errors, (c) an emphasis on the importance of using the scale definitions and anchors to make the ratings, and (d) a within-ratee card sorting exercise to prevent intra-ratee halo error. The sorting exercise required each rater to read cards showing the rating scales and, for each peer ratee, to sort the cards into three piles according to areas that were (a) strong, (b) adequate, or (c) in need of improvement for the ratee. After having completed ratings of current performance, the Soldiers received an oral briefing from the administrator describing the conditions under which Soldiers will need to perform in the future. After the briefing, Soldiers received instructions for completing the Future Expected Performance Rating Scales.

Throughout the rating process, Soldiers were monitored by administrators to ensure that forms were completed correctly.

Supervisor Sessions

The role of supervisors was to rate the job performance of participating Soldiers. A supervisor was eligible for participation if he or she had known one or more participating Soldiers for at least one month. In addition to the project briefing and Supervisor Background Information Form, in-processing for each supervisor included completing a rating card listing the names and identification numbers for each Soldier being rated. Supervisors were encouraged to rate as many as 10 of the participating Soldiers. The structure and content of the supervisor performance rating session was the same as the performance rating portion of the Soldier paper-and-pencil session.

Experience has shown the difficulty of getting supervisor ratings for every participating Soldier during a site visit (Keenan et al., 2005; Knapp, McCloy, & Heffner, 2004). Anticipating this challenge, we asked each Soldier to identify two supervisor raters who could provide performance ratings. A "mail-back" procedure was developed for the identified Supervisors of any Soldier who was not rated by at least one supervisor during the site visit. At the end of the site visit, arrangements were made for delivery of a self-administered "mail-back" packet to these supervisors. Each packet included a description of the project, instructions for completing the ratings, future Army conditions briefing slides with notes, relevant rating scales and answer sheets, and return envelopes. The card-sorting exercise was not included in the mail-back packets. Of the 388 participating supervisors referred to in Table 2.1, 79 did so via the mail-back procedure.

Staff Training

HumRRO and ARI personnel served as test administrators. Separate test administration manuals were developed for the Soldier and supervisor sessions. These manuals included sections containing the following information:

- Session schedules (i.e., timing and order of administration)
- Instructions for preparing Soldier and supervisor packets containing forms to be completed by participants (separate packets for AW, 11B, and 25U Soldier and supervisor participants)
- Instructions for setting up computer and paper-and-pencil Soldier rooms and supervisor rooms
- Instructions for in-processing participants (e.g., determining eligibility of participants, project briefings, and background information forms)
- Instructions for administering sessions
- Procedures for data documentation and quality control (e.g., storing data collected on computers and by paper-and-pencil, checking data, and preparing supervisor mail-back packets)
- Procedures for sending equipment and data back to HumRRO

In addition to reviewing manuals, data collectors participated in a half-day training session before the site visits. The training reviewed and supplemented the materials in the test administration manuals.

Database Construction

Several procedures were implemented to maximize the completeness and quality of the data collected. Beyond the manuals and administrator training, data collection logs were kept to record relevant events that occurred during each session. Log entries included information on such occurrences as (a) environmental events that might affect the quality of the data (e.g., loud construction next door); (b) Soldiers who were observed to be inattentive, pattern responding, or just not following instructions; and (c) computer malfunctions during testing. This section covers the initial processing and scrubbing of data, the addition of archival data from Army records, and data cleaning and imputation.

Initial Processing and Scrubbing

Four major types of data had to be processed and combined: (a) Soldier responses on scannable forms collected during paper-and-pencil sessions, (b) Soldier responses collected electronically during computer sessions, (c) peer and supervisor performance ratings collected on scannable forms during Soldier paper-and-pencil and supervisor sessions, and (d) archival data collected from Army records.

After the computer data had been integrated with data from the scannable forms, data were examined for logical inconsistencies. Examples of observed anomalies included (a) two sets of responses on a test for a single participant, (b) missing computer data for a participant on a single test, or (c) illogical responses on rating scannable forms. As described below, the database manager used the session logs and various data analysis techniques to resolve as many of these anomalies as possible.

Soldier data on demographic (e.g., gender, race/ethnicity, and start date) and other variables (e.g., ASVAB test scores) were retrieved from the Enlisted Master File (EMF) and Military Enlistment Processing Command Integrated Resource System (MIRS). These data were accessed by matching the SSNs of Soldiers in the Select21 database with Soldier SSNs in the archival databases.

Data Cleaning and Imputation

After the initial data were processed and prepared by the database manager, data analysts conducted additional cleaning and imputation analyses. The session logs and analyses examining different types of pattern responding were used to identify Soldiers and supervisors with questionable data that should be dropped. A Soldier's or supervisor’s responses for a particular instrument were dropped if the participant failed to respond to at least 90% of the items. Data from Soldiers who completed computerized instruments too quickly were also dropped. Finally, in an effort to achieve the largest possible sample sizes for criterion-related validity analyses, missing responses on instruments were imputed where possible. One imputation method was a multiple-regression based strategy that used responses to other items to impute the missing response to a given item. Another

approach used for self-report instruments with multiple items per scale was to use the mean score on the items to which the Soldier responded as the scale score, as long as the participant responded to enough items on the scale to provide a sufficiently reliable score. This simpler approach is sufficient when scale scores are used for subsequent analyses and item scores are not. Finally, for Soldiers with missing self-report data on gender and race/ethnicity, these values were imputed using data from the EMF archival database. Additional data cleaning, imputation, and scoring details are provided in the individual instrument chapters. Instrument scale and composite scores are included in the final database along with item-level data.

Sample Sizes

Table 2.4 shows samples sizes by important demographic variables. The sample sizes for individual instruments vary based on instrument-specific data cleaning and imputation analyses. In the remaining chapters, subgroup difference and differential prediction analyses are presented for gender, race, ethnicity, and MOS cluster. As mentioned in Chapter 1, the number of 25U Soldiers was not sufficient to treat the 25U and the 11B Soldiers as separate samples with MOS-specific criterion measures. However, we were able to organize 710 of the 812 participating Soldiers into MOS clusters. MOS cluster membership served as a subgrouping variable in subsequent analyses designed to give a sense of the classification potential of the Select21 predictors (see Chapter 14). The MOS clusters were derived from the Select21 future-oriented job analysis (Sager, Russell, Campbell, & Ford, 2005); Table 2.5 shows brief definitions of each of the four clusters for which we had sufficient numbers of Soldiers to analyze as subgroups.

Table 2.4. Select21 Concurrent Validation Sample Sizes by Subgroup

	Wave 1		Wave 2		Full Sample	
Subgroup	*n*	%	*n*	%	*N*	%
Gender						
Male	572	91.1	156	85.2	728	89.8
Female	56	8.9	27	14.8	83	10.2
Race						
White	401	63.8	119	65.0	520	64.0
Black	127	20.2	35	19.1	162	20.0
Other	101	16.1	29	15.8	130	16.0
Ethnicity						
White Non-Hispanic	350	55.7	105	57.4	455	56.1
Hispanic	120	19.1	36	19.7	156	19.2
MOS Cluster						
Close Combat	297	54.6	85	52.8	382	53.8
SINC	64	11.8	50	31.1	114	16.1
Maintenance/Repair	99	18.2	9	5.6	121	17.0
Logistics/Supply	84	15.4	17	10.6	93	13.1
MOS Sample						
Army-Wide	448	71.2	91	49.7	539	66.4
11B Infantryman	131	20.8	85	46.4	216	26.6
25U Signal Support Systems Specialist	50	7.9	7	3.8	57	7.0
Total	629		183		812	

Note. SINC = Surveillance, Intelligence, and Communications. % = Percentage within sample.

Awareness of the small demographic differences between Wave 1 and Wave 2 (see Table 2.5) and the somewhat greater differences between the full data set and the Army enlisted population (Office of the Under Secretary of Defense, Personnel and Readiness, 2004) could be useful when interpreting some results in the remaining chapters. Wave 2 had a slightly greater percentage of females than Wave 1. Relative to currently enlisted Army personnel (15% female), the full data set had a smaller percentage of females. Waves 1 and 2 showed similar representation by race; however, the full data set had a relatively smaller percentage of Black Soldiers compared to the current Army (25% for current enlisted Army personnel). Waves 1 and 2 also had similar representation by ethnicity; however, the full data set had a relatively higher percentage of Hispanic Soldiers than the current Army (11% for current enlisted Army personnel). Across MOS clusters, Waves 1 and 2 were similar in terms of Close Combat (CC) representation; however, Wave 1 had relatively fewer Surveillance, Intelligence, and Communications (SINC) Soldiers and Wave 2 had relatively fewer Maintenance/Repair and Logistics/Supply Soldiers. The *Population Representation in the Military Services* report (Office of the Under Secretary of Defense, Personnel and Readiness, 2004) does not organize its data according to our MOS clusters, so comparison is difficult. However, it does suggest that the Select21 concurrent validation full data set had a much higher percentage of CC Soldiers than the current enlisted Army population.

Table 2.5. MOS Cluster Definitions

Close Combat	MOS in this cluster emphasize (a) closing with and destroying enemy personnel, weapons, equipment, and structures, using fire maneuver, in both offensive and defensive operations; and (b) controlling, denying, or occupying disputed or hostile terrain.
Surveillance, Intelligence, and Communications	MOS in this cluster provide (a) surveillance; (b) intelligence; and (c) video, voice, and data communications support to forces in tactical environments. This includes information about the location and disposition of the enemy and facilitation of communications among friendly forces.
Maintenance	MOS in this cluster required Soldiers to install, repair, and maintain mechanical, electronic, and aviation equipment. Activities include inspection, damage assessment, use of diagnostic instruments, and troubleshooting. This cluster is based on a combination of three original Select21 job analysis clusters covering mechanical, electronic, and aircraft repair, respectively.
Logistics/Supply	MOS in this cluster focus on providing support to deployed troops. Activities include (a) operating transportation vehicles, (b) preparing supplies for shipment, (c) unloading and unpacking supplies, (d) maintaining inventory records, and (e) distributing supplies.

Note. From *Future Soldiers: Analysis of Entry-Level Performance Requirements and Their Predictors* (Technical Report 1169; p. C-1 to C-8), by C.E. Sager, T.R. Russell, R.C. Campbell, and L.A. Ford, Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. Summarized with permission.

Summary

This chapter described the Select21 concurrent validation data collection effort and procedures for processing and cleaning the data. Participants included 812 first-term enlisted Soldiers and 388 of their supervisors. Soldiers completed a number of experimental criterion and predictor measures using laptop computers and paper-and-pencil forms. Criteria included measures of job knowledge, job satisfaction, and supervisor and peer ratings of observed and future expected job performance. Predictors included measures of psychomotor ability, judgment, interests and values, and temperament constructs hypothesized to be relevant to the performance of first-term Soldiers. The remaining chapters present and discuss analyses addressing the psychometric characteristics of the experimental criterion and predictor measures and the criterion-related validity of the predictors.

PART 2: VALIDATION CRITERIA

CHAPTER 3: ATTITUDINAL CRITERION MEASURES

Dan J. Putka and Chad H. Van Iddekinge
HumRRO[4]

Overview

In addition to job performance, two criteria of interest to the Army for evaluating the efficacy of experimental selection measures are attrition and re-enlistment behavior. To fully investigate such criteria, a longitudinal research design is needed. Given the concurrent nature of the Select21 validation effort, it was not possible to examine attrition and re-enlistment outcomes. Therefore, two criterion measures, the Army Life Survey and the Future Army Life Survey, were developed to assess the attitudinal pre-cursors of attrition and re-enlistment behavior. The scales that comprise these measures reflect constructs that theory and empirical evidence suggest are the strongest precursors of attrition and re-enlistment (e.g., Ajzen, 1991; Hom & Griffith, 1995; Strickland, 2005). The constructs assessed in the aforementioned measures reflect both current-state and future-oriented criteria. Current-state criteria reflect Soldiers' current standing on a construct (e.g., current level of job satisfaction), whereas future-oriented criteria reflect Soldiers' expected future standing on a construct given anticipated future Army conditions.

Instrument Descriptions

Army Life Survey

Current-state criteria are assessed in the Army Life Survey (ALS). The ALS is a 99-item instrument comprising 15 scales. These scales were developed based on a review of research from the applied psychology literature (e.g., Hom & Griffeth, 1995; Jex, 1998; Meyer & Allen, 1991; Spector, 1997) and previous Army research, such as Project A (Campbell & Knapp, 2001) and Project First Term (Strickland, 2005). In fact, most of the ALS scales were adapted from established measures within the literature. Details on the development of these scales were presented in Van Iddekinge, Putka, and Sager (2005). To score the ALS, items for each scale described in Table 3.1 were averaged together to create a total score for that scale.

The 15 scales on the ALS can be grouped into two broad categories of criterion constructs. The first category includes two constructs believed to be most proximal to Soldiers' choice to remain in the Army, namely attrition cognitions and career intentions (Strickland, 2005). The second category of ALS constructs includes measures of several attitudinal variables that have been shown to underlie both intentions to leave and actual withdrawal behavior (e.g., Griffeth, Hom, & Gaertner, 2000; Strickland, 2005). These include satisfaction with various aspects of Army life, organizational commitment, perceived fit, perceived stress, and perceived importance of the seven "Core Army Values." Although more distal to attrition and re-enlistment behavior compared to attrition cognitions and career intentions, these attitudinal variables are expected to be more proximal to the Select21 predictors, in particular the person-environment fit

[4] Chad H. Van Iddekinge is currently an assistant professor in the College of Business at Florida State University.

predictors discussed in Chapters 10 and 11. This is not to imply that we expected person-environment fit predictors would be unrelated to attrition and re-enlistment (and their intention-related precursors). Rather, we expected that they would be more strongly related to attitudinal variables such as job satisfaction and organizational commitment, given that such predictors are more proximal to attitudes in the causal chain hypothesized to link P-E fit predictors to behavioral outcomes such as attrition and re-enlistment.

Table 3.1. ALS Scale Descriptions

Scale	Description
Satisfaction with Supervision	Five-item scale assessing Soldiers' satisfaction with the supervision they receive.
Satisfaction with Peers	Four-item scale assessing Soldiers' satisfaction with their co-workers.
Satisfaction with Work Itself	Seven-item scale assessing Soldiers' satisfaction with working in their MOS.
Satisfaction with Promotions	Four-item scale assessing Soldiers' satisfaction with their promotions.
Satisfaction with Pay and Benefits	Five-item scale assessing Soldiers' satisfaction with their pay and benefits.
Satisfaction with the Army	Ten-item scale assessing Soldiers' satisfaction with Army life in general.
Affective Commitment	Eight-item scale assessing Soldiers' feelings of wanting to remain in the Army.
Continuance Commitment	Seven-item scale assessing Soldiers' feelings of needing to remain in the Army.
Normative Commitment	Five-item scale assessing Soldiers' feelings of obligation to remain in the Army.
Perceived MOS Fit	Six-item scale assessing how well Soldiers perceive themselves fitting in their MOS.
Perceived Army Fit	Six-item scale assessing how well Soldiers perceive themselves fitting in the Army in general.
Perceived Stress	Nine-item scale assessing Soldiers' perceived level of stress.
Attrition Cognitions	Three-item scale assessing the degree to which Soldiers have thoughts of attriting.
Career Intentions	Five-item scale assessing Soldiers' intentions to re-enlist and make the Army a career.
Core Army Values	Seven-item scale assessing the extent to which Soldiers perceive "Core Army Values" as important.

Future Army Life Survey

The Future Army Life Survey (FALS) is a 29-item measure that assesses Soldiers' attitudes and perceptions of work conditions that are expected to become more common in the Army as it transforms to the Future Force. The FALS was designed to assess several of the general attitudinal constructs measured in the ALS. Specifically, the FALS measures (a) expected attachment to or

liking of the Army under future conditions (Future Army Affect), (b) perceived stressfulness of future conditions (Future Stress), and (c) expected performance under future conditions (Future Performance Efficacy). To give Soldiers a context for responding to the FALS, they were asked to read descriptions of anticipated future Army conditions (e.g., frequent change, continuous learning) prior to completing the survey. These conditions were based on the Select21 future-oriented job analysis (Sager, Russell, Campbell, & Ford, 2005). The three scales that comprise the FALS are shown in Table 3.2. Details on the development of these scales were presented in Van Iddekinge, Putka et al. (2005).[5] Items for each scale were averaged together to create a total score for that scale.

Table 3.2. FALS Scale Descriptions

Scale	Description
Future Performance Efficacy	Seven-item scale assessing Soldiers' perceived ability to perform well under expected future Army conditions.
Future Stress	Five-item scale assessing the extent to which Soldiers perceive expected future Army conditions to be stressful.
Future Army Affect	Five-item scale assessing the extent to which Soldiers have positive feelings about expected future Army conditions.

Psychometric Properties of the Attitudinal Criteria

A total of 786 Soldiers completed the ALS, and 772 Soldiers completed the FALS during the concurrent validation data collections.[6] We did, however, eliminate the responses of 46 Soldiers who test administrators flagged as having questionable ALS data or who had exhibited extremely unlikely patterns of responding (mostly the latter), and responses of 52 Soldiers with similarly questionable data on the FALS. Thus, the analysis sample comprised 740 Soldiers for the ALS and 720 Soldiers for the FALS.

Descriptive Statistics and Reliability

Table 3.3 shows descriptive statistics and internal consistency reliability estimates for the ALS and FALS scales. Estimates were computed by sample (e.g., Wave 1, Wave 2) to facilitate validation work reported in subsequent chapters. With the potential exception of ALS Attrition Cognitions (Full Sample $\alpha = .68$), the ALS and FALS scales exhibited good levels of internal consistency (i.e., α's > .75) and variability.

[5] Note that the FALS scale names used in this chapter are different from those used in the measure development report (Van Iddekinge, Putka et al., 2005). This was done to reflect adjustments made to scale content after completion of the criterion field test. In the earlier report, the Future Army Fit scale (now named Future Performance Efficacy) reflected a mix of satisfaction and performance-related items. For purposes of the concurrent validation, we eliminated six items that tapped satisfaction to make this measure more distinct from the Future Army Affect scale (formerly named Future Continuance). Two items were also dropped for the Future Army Affect scale (those reflecting relative comparisons with the current Army). We renamed both of these scales because we felt that the new names provided more accurate descriptions of the scale content. Comparison of results presented later in this chapter to those presented in Van Iddekinge, Putka et al. (2005) reveals that dropping the aforementioned items had minimal impact had on the psychometric quality (e.g., reliability, variability) of these scales.

[6] Information on the demographic characteristics of Soldiers who completed the measures discussed in this chapter is provided in Chapter 2.

Table 3.3. ALS and FALS Descriptive Statistics and Reliability Estimates by Sample

	Sample								
	Wave 1			Wave 2			Full Sample		
Instrument/Scale	α	*M*	*SD*	α	*M*	*SD*	α	*M*	*SD*
ALS									
Satisfaction with Supervision	0.88	3.06	0.87	0.91	3.29	0.95	0.89	3.12	0.89
Satisfaction with Peers	0.84	3.62	0.78	0.81	3.62	0.71	0.83	3.62	0.77
Satisfaction with Work Itself	0.91	2.98	0.90	0.92	3.07	0.96	0.91	3.00	0.91
Satisfaction with Promotions	0.88	2.92	0.99	0.92	3.07	1.09	0.89	2.96	1.01
Satisfaction with Pay and Benefits	0.90	2.72	0.93	0.91	2.74	1.00	0.90	2.72	0.95
Satisfaction with the Army	0.87	2.90	0.76	0.89	3.03	0.83	0.87	2.93	0.78
Affective Commitment	0.89	2.81	0.87	0.91	2.91	0.97	0.89	2.83	0.89
Continuance Commitment	0.87	2.39	0.91	0.89	2.49	1.01	0.88	2.41	0.94
Normative Commitment	0.84	2.02	0.87	0.89	2.15	1.00	0.86	2.05	0.91
Perceived MOS Fit	0.85	3.00	0.91	0.91	3.08	1.04	0.87	3.02	0.94
Perceived Army Fit	0.79	3.05	0.80	0.83	3.20	0.87	0.80	3.08	0.82
Perceived Stress	0.76	3.23	0.64	0.78	3.10	0.67	0.76	3.20	0.65
Attrition Cognitions	0.67	2.25	0.97	0.71	2.14	0.99	0.68	2.22	0.98
Career Intentions	0.93	1.99	1.08	0.95	2.24	1.18	0.93	2.05	1.11
Core Army Values	0.93	4.12	0.89	0.94	4.26	0.85	0.94	4.16	0.88
FALS									
Future Performance Efficacy	0.86	3.63	0.67	0.92	3.71	0.84	0.88	3.65	0.72
Future Stress	0.77	3.03	0.71	0.78	2.95	0.74	0.77	3.01	0.71
Future Army Affect	0.87	3.11	0.89	0.93	3.13	1.07	0.89	3.11	0.93

Note. $n_{\text{Wave 1}}$ = 505-564. n_{Wave2} = 173-176. $n_{\text{Full Sample}}$ = 680-740. Reliability estimates are Cronbach's alphas.

Scale Intercorrelations

Table 3.4 shows raw zero-order intercorrelations among the ALS and FALS scales. Given the similarity of constructs assessed by some of the ALS scales, we were concerned about the potential for overly high relations among scale scores. However, the scale correlations suggest that this concern is not a significant issue. Intercorrelations ranged from -.68 (Perceived Stress and Perceived Army Fit) to .78 (Affective Commitment and Perceived Army Fit). The mean absolute correlation among the 15 ALS scales was .36.

The FALS scales were moderately intercorrelated, with Future Performance Efficacy and Future Army Affect being the most strongly correlated (r = .59). In creating the FALS, one concern was whether its scales would be distinct from the ALS scales. Specifically, despite the effort to reduce halo between concurrent and future measures, much of the variance in FALS scales could simply reflect Soldiers' current attitudes towards the Army. The correlations presented in Table 3.4 inform this question. On average, relations between ALS and FALS scales were small to moderate (average absolute r = .25), with no correlation exceeding .53 in magnitude. The strongest ALS correlates of Future Performance Efficacy were current Perceived Army Fit (r = .42), Core Army Values (r = .41), and Perceived Stress (r = -.40). The strongest ALS correlates of Future Army Affect were current Perceived Army Fit (r = .53), Affective

Table 3.4. ALS and FALS Scale Intercorrelations

	ALS															FALS	
	Satisfaction Scales						Commitment Scales			Fit Scales		Other Scales on the ALS					
Instrument/Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ALS																	
1. Satisfaction with Supervision																	
2. Satisfaction with Peers	.33																
3. Satisfaction with Work Itself	.57	.37															
4. Satisfaction with Promotions	.53	.29	.50														
5. Satisfaction with Pay and Benefits	.33	.24	.34	.42													
6. Satisfaction with the Army	.59	.41	.60	.61	.52												
7. Affective Commitment	.41	.35	.48	.43	.34	.68											
8. Continuance Commitment	.28	.11	.34	.23	.23	.44	.58										
9. Normative Commitment	.38	.18	.45	.35	.26	.52	.68	.67									
10. Perceived MOS Fit	.37	.27	.51	.36	.27	.48	.46	.25	.34								
11. Perceived Army Fit	.46	.38	.49	.45	.35	.71	.78	.49	.55	.53							
12. Perceived Stress	-.49	-.29	-.46	-.43	-.36	-.66	-.57	-.36	-.46	-.43	-.68						
13. Attrition Cognitions	-.37	-.27	-.31	-.37	-.22	-.49	-.50	-.28	-.33	-.33	-.55	.49					
14. Career Intentions	.33	.19	.38	.35	.25	.50	.56	.60	.65	.36	.57	-.48	-.36				
15. Core Army Values	.20	.28	.25	.24	.13	.38	.48	.19	.25	.29	.50	-.34	-.34	.25			
FALS																	
16. Future Performance Efficacy	.21	.21	.17	.26	.08	.35	.36	.12	.21	.29	.42	-.40	-.33	.33	.41		
17. Future Stress	-.07	-.11	**-.05**	-.13	-.07	-.15	-.11	**-.04**	-.09	-.07	-.17	.29	.15	-.13	-.14	-.36	
18. Future Army Affect	.27	.19	.32	.30	.13	.45	.53	.38	.44	.32	.53	-.42	-.30	.50	.33	.59	-.27

Note. n = 688-740. All correlations in this table are raw zero-order correlations. All correlations are statistically significant ($p < .05$, one-tailed), except those that are bolded.

Commitment (r = .53), and Career Intentions (r = .50). Of the three FALS scales, Future Stress was the least related to the ALS scales. The strongest ALS correlate of Future Stress was current Perceived Stress (r = .29). Thus, at the bivariate level, the FALS scales appeared to be related to, yet distinct from, Soldiers' attitudes towards the current Army. These findings are consistent with results from the criterion field test (Van Iddekinge, Putka et al., 2005). These findings also suggest that that Soldiers' perceptions of the future Army were not simply a function of their attitudes towards the current Army.

Subgroup Differences

Tables 3.5 and 3.6 show subgroup means on ALS and FALS scales by gender and race/ethnic group. In terms of gender, there were only five statistically significant mean differences, and the effect sizes associated with those differences were modest. Specifically, female Soldiers had mean scores on Attrition Cognitions that were 0.34 *SD*s higher than scores for male Soldiers, whereas males had mean scores that were 0.25 to 0.32 *SD*s higher than scores for females on Satisfaction with Peers, Satisfaction with the Army, Future Performance Efficacy, and Future Army Affect. Similarly small differences were found on the ALS and FALS scales across race/ethnic groups. Though some differences were statistically significant, the magnitudes of their effects were modest. For example, the largest difference found between White and Black Soldiers was on ALS Attrition Cognitions, with Blacks having scores that were 0.37 *SD*s higher than Whites.

Table 3.5. ALS and FALS Scale Scores by Gender

		Male		Female	
Instrument/Scale	d_{FM}	*M*	*SD*	*M*	*SD*
ALS					
Satisfaction with Supervision	-0.23	3.14	0.89	2.93	0.89
Satisfaction with Peers	**-0.29**	3.64	0.77	3.42	0.70
Satisfaction with Work Itself	-0.01	3.00	0.92	2.99	0.90
Satisfaction with Promotions	-0.11	2.97	1.02	2.86	0.97
Satisfaction with Pay and Benefits	-0.08	2.73	0.95	2.66	0.92
Satisfaction with the Army	**-0.25**	2.96	0.78	2.76	0.72
Affective Commitment	-0.13	2.84	0.90	2.73	0.83
Continuance Commitment	0.02	2.41	0.94	2.42	0.94
Normative Commitment	-0.06	2.06	0.92	2.00	0.80
Perceived MOS Fit	-0.12	3.03	0.94	2.92	0.92
Perceived Army Fit	-0.04	3.09	0.82	3.05	0.84
Perceived Stress	-0.08	3.20	0.65	3.15	0.64
Attrition Cognitions	**0.34**	2.19	0.97	2.52	1.03
Career Intentions	0.00	2.05	1.10	2.05	1.16
Core Army Values	-0.05	4.16	0.89	4.12	0.82
FALS					
Future Performance Efficacy	**-0.32**	3.67	0.73	3.44	0.58
Future Stress	0.01	3.01	0.72	3.02	0.70
Future Army Affect	**-0.29**	3.14	0.94	2.87	0.89

Note. n_{Male} = 641-659, n_{Female} = 78-80. d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of non-referent group – mean of referent group)/*SD* of referent group. Referent groups (e.g., Males) are listed second in the effect size subscript. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.6. ALS and FALS Scale Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
Instrument/Scale	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ALS										
Satisfaction with Supervision	-0.03	0.01	3.10	0.87	3.07	1.01	3.12	0.85	3.12	0.90
Satisfaction with Peers	0.06	0.10	3.61	0.75	3.66	0.85	3.59	0.76	3.66	0.71
Satisfaction with Work Itself	0.08	0.18	2.97	0.90	3.04	0.97	2.95	0.91	3.11	0.89
Satisfaction with Promotions	-0.10	0.10	2.99	1.01	2.89	1.02	2.97	0.98	3.07	1.06
Satisfaction with Pay and Benefits	0.10	**0.24**	2.69	0.94	2.79	1.00	2.66	0.94	2.89	0.93
Satisfaction with the Army	-0.07	0.13	2.94	0.78	2.89	0.80	2.93	0.77	3.03	0.78
Affective Commitment	**-0.28**	0.11	2.89	0.90	2.63	0.87	2.87	0.90	2.97	0.86
Continuance Commitment	0.10	0.17	2.38	0.94	2.47	0.96	2.35	0.92	2.51	0.98
Normative Commitment	0.01	0.13	2.04	0.93	2.05	0.85	2.02	0.93	2.15	0.89
Perceived MOS Fit	-0.16	0.03	3.06	0.96	2.91	0.91	3.05	0.98	3.08	0.87
Perceived Army Fit	-0.16	0.12	3.10	0.82	2.97	0.80	3.09	0.83	3.19	0.77
Perceived Stress	-0.01	-0.09	3.20	0.65	3.19	0.67	3.21	0.66	3.15	0.60
Attrition Cognitions	**0.37**	0.02	2.15	0.98	2.51	0.97	2.15	0.98	2.17	0.93
Career Intentions	0.08	0.02	2.03	1.12	2.12	1.12	2.03	1.15	2.05	1.03
Core Army Values	**-0.36**	0.00	4.22	0.85	3.91	0.94	4.22	0.85	4.22	0.85
FALS										
Future Performance Efficacy	**-0.31**	-0.04	3.69	0.71	3.47	0.73	3.69	0.71	3.67	0.70
Future Stress	-0.14	-0.13	3.02	0.72	2.93	0.70	3.05	0.72	2.95	0.73
Future Army Affect	-0.18	**0.21**	3.12	0.92	2.95	0.97	3.08	0.93	3.28	0.87

Note. n_{White} = 523-533, n_{Black} = 134-135. $n_{White\ Non\text{-}Hispanic}$ = 411-417, $n_{Hispanic}$ = 134-144. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of non-referent group – mean of referent group)/*SD* of referent group. Referent groups (e.g., Whites, White Non-Hispanics) are listed second in the effect size subscript. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Reducing the Number of Attitudinal Criteria

Given the large number of criterion scores, as well as the conceptual overlap between many of the ALS and FALS scales, we reduced the number of attitudinal criteria for use in validating predictors in subsequent chapters. We began by factor analyzing the 15 ALS scales. This analysis was conducted using principal axis factoring with oblique rotation. It revealed a three-factor solution that accounted for 62.8% of the variance among the ALS scales. The first factor reflected a "general satisfaction" factor, and comprised the six ALS satisfaction scales, Perceived MOS Fit, and Perceived Stress (negative loading). The second factor reflected a "commitment to remain" factor, and comprised Continuance Commitment, Normative Commitment, and Career Intentions. Lastly, the third factor reflected a "positive Army affect" factor, and included Perceived Army Fit, Core Army Values, Affective Commitment, and Attrition Cognitions (negative loading).[7] Given the heterogeneous nature of the aforementioned factors, as well as the zero-order correlations presented in Table 3.4 (which indicate the ALS scales have a substantial portion of unique variance specific to their targeted construct), we were not comfortable with aggregating scales within factors to create composite scores. Creating an aggregate score based on the first factor, for example, would have resulted in the combination of constructs (e.g., stress and satisfaction) that have been clearly differentiated in the research literature.

Accordingly, we took a rational approach to select a subset of the ALS and FALS scales to focus on in subsequent predictor validation analyses. In deciding which scales to move forward with, we took into account several factors. At a basic level, one consideration was that the selected scales should be fairly general (i.e., they should cover a lot of the criterion space of interest), yet at the same time be conceptually distinct from the other scales chosen. We also favored scales with particularly good psychometric properties (e.g., reliability, score variability).

Another consideration was the strength of the relation we expected between the Select21 predictors and the given ALS/FALS scale. Past research has suggested that the strongest attitudinal correlates of the Select21 interests and values measures (see Chapters 10 and 11) would be the satisfaction and perceived fit scales (e.g., Dawis & Lofquist, 1984; Holland, 1985; Kristof, 1996). In other words, these would be the criteria that we would expect to have the strongest relationship with the Select21 predictors in the hypothesized causal chain linking the predictor space to attrition and re-enlistment behavior.

A final consideration was the strength of relation we expected between the predictors and the ultimate criteria of interest, in this case attrition and re-enlistment behavior. As noted earlier, past research has suggested scales measuring attrition cognitions and career intentions should be the strongest predictors of actual attrition and re-enlistment behavior (Strickland, 2005).

Based on these considerations, we chose five scales on which to focus for the validation effort summarized in the remainder of this report. Four of the five scales were drawn from the ALS, namely: (a) Satisfaction with the Army, (b) Perceived Army Fit, (c) Attrition Cognitions, and (d) Career Intentions; and one scale was drawn from the FALS, namely, Future Army

[7] Affective Commitment also cross-loaded highly on the second factor.

Affect.[8] The choice of these five scales is desirable on several fronts. First, it includes both current and future-oriented constructs. Second, it strikes a balance in terms of proximity of the chosen scales to the Select21 predictors and actual attrition and re-enlistment behavior. Lastly, the chosen scales are relatively easy to understand and explain to Army decision-makers.

Summary

Overall, the psychometric properties of the ALS and FALS scales are good. All scales exhibited sufficient levels of variance and had acceptable levels of internal consistency. Correlations among scales were moderate, suggesting that scales were not overly redundant with one another. The FALS scales exhibited only small to moderate correlations with ALS scales, indicating that Soldiers' attitudes toward the future Army are not simply a function of their attitudes about the current Army. Lastly, although many ALS and FALS scales were examined in this chapter, given the conceptual overlap among them, for sake of parsimony we chose five of them for use in subsequent validation analyses. The scales chosen were Satisfaction with the Army in General, Perceived Army Fit, Attrition Cognitions, Career Intentions, and Future Army Affect. These scales were chosen based on empirical (e.g., factor analyses) and rational considerations (e.g., hypothesized strength of relation to predictors and attrition and re-enlistment criteria).

[8] Although the internal consistency reliability of the Attrition Cognitions was somewhat lower than other scales considered for inclusion in this final set of criteria, it is important to remember that corrections for attenuation due to measurement error will be made when reporting criterion-related validity estimates in later chapters of this report.

CHAPTER 4: BASIC PERFORMANCE CRITERION SCORES

Huy Le, Patricia Keenan, Gordon Waugh, Maggie Collins, and Dan Putka
HumRRO

Overview

We developed four criterion measures that were intended, as a set, to provide reasonably comprehensive coverage of the 19 Army-wide performance requirements for first-tour Soldiers identified in the job analysis (Sager, Russell, Campbell, & Ford, 2005). These measures include performance rating scales (covering both current observed performance and expected future performance), a job knowledge test, a situational judgment test, and a self-report measure of awards, educational experiences, and disciplinary actions.

This chapter summarizes the criterion instruments and describes their psychometric properties (i.e., descriptive statistics, reliabilities, and intercorrelations). Complete descriptions of all criterion measures can be found in the Select21 measure development report (Knapp, Sager, & Tremble, 2005). Chapter 5 of the present report provides information about higher-order performance composites.

Instrument Descriptions and Scoring

Performance Rating Scales

Supervisors and peers rated Soldiers' current performance using the Army-Wide Current Observed Performance Ratings Scales (AW COPRS) and then used the Army-Wide Future Expected Rating Scales (AW FX) to rate those Soldiers' expected performance under future conditions.

Army-Wide Current Observed Performance Ratings Scales (AW COPRS)

The AW COPRS contain rating scales for 12 performance dimensions, such as "Supports Peers" and "Exhibits Effort and Initiative on the Job." The instrument also includes a single overall performance effectiveness scale.

Prior to making ratings, raters received training on the format of the scales and how to use them accurately. The training focused on the importance of reading the anchors, thinking about a Soldier's relative strengths and weaknesses, and applying that insight to the ratings. To reinforce this idea, training included a performance dimension sorting task designed to familiarize raters with the dimension definitions and assist them in identifying the relative strengths and weaknesses of each ratee. Raters sorted the cards into three categories – "Needs Improvement," "Adequate," and "Strong" – to reflect the performance level of the Soldier they were rating. Training also included admonitions about response tendencies (e.g., halo error) and evaluation biases and stressed the notion that the ratings would be used for research purposes only, to lessen the tendency of raters to "help" their subordinate or buddy by providing lenient ratings.

Army-Wide Future Expected Performance Rating Scales (AW FX)

These scales yield ratings of expected Soldier effectiveness under projected future conditions. The AW FX addresses the following four future conditions identified during the job analysis: (a) learning environment, (b) disciplined initiative, (c) communication method and frequency, and (d) individual pace and intensity.

Prior to making their AW FX ratings, raters received a briefing that described the most important changes expected to occur in the Army of the future. The FX rating booklet provided additional information about the four future conditions and a rating scale for each condition. Descriptions for all the future conditions may be found in the Select21 measure development report (Appendix E, Knapp et al., 2005). Raters used a 7-point rating scale to make an overall effectiveness rating for each condition.

Scoring the Performance Ratings

We computed the average rating across raters (peer and supervisor) for each dimension in the AW COPRS (e.g., Common Task Performance) and each condition in the AW FX scales (e.g., Learning Environment) to derive scale scores. Chapter 5 describes how the scale-level ratings contributed to the performance model.

Army-Wide Job Knowledge Test (AWJKT)

The job knowledge test is a "can-do" measure of first-term Soldier performance, designed to measure Soldiers' knowledge of common tasks (e.g., land navigation, first aid, survival). Select21 test developers used a variety of item formats (e.g., multiple-choice, drag and drop, ranking, matching) and graphics to enhance the realism of these computer-administered tests and minimize reading requirements.

Project staff drafted the tests using blueprints developed from the Select21 job analysis results and subject matter expert (SME) input. The test blueprints (i.e., content specifications, including the degree to which each content area is reflected in the test) are composed of the performance requirements that could reasonably be assessed in a written test (e.g., knowledge of first aid procedures is more easily tested than oral communication skill by this method).

Scoring the AWJKT

This section provides a brief description of how the different types of items were scored. More detailed descriptions can be found in the measure development report (Collins, Le, & Schantz, 2005).

Multiple-choice item analyses. We assigned a score of 1 for a correct response and zero for an incorrect response to a multiple-choice item. We used classical item statistics to analyze these questions. These statistics include the percentage of examinees selecting each response option and the point-biserial correlation between the option selected and total score of all the multiple-choice items.

Non-traditional item analyses. The non-traditional items (e.g., matching, ranking) allow scoring options that are not dichotomous so that examinees can get partial credit for getting some (but not all) parts of the items correct. Adopting partial-credit scoring procedures for non-traditional items, however, can result in assigning more weight to these items in the total score (as compared to the traditional multiple-choice items) than desired (Wainer & Thissen, 2001). Thus, we used analytic procedures described in detail in Collins et al. (2005) to score and then weight the non-traditional items. Optimal weighting serves two purposes: (a) ensuring that items are combined most efficiently to minimize the effect of measurement error, and (b) providing a benchmark for the non-traditional items (against the multiple-choice items) that facilitated final selection of items in accordance with the test blueprints. A composite score was calculated that summed the selected multiple-choice and appropriately weighted non-traditional items. This score is reported as a percentage correct score to facilitate its interpretability.

Criterion Situational Judgment Test (CSJT)

The CSJT is a 27-item situational judgment test. Each item consists of the description of a problem situation (i.e., scenario) followed by four actions that a Soldier might take in that situation. The scenarios and response options were written by NCOs in a series of workshops. The scenarios represent situations that Soldiers with 18–36 months of experience might encounter. They were developed to tap the following performance dimensions: Adapts to Changing Situations, Relates to and Supports Peers, Exhibits Self-Management, Exhibits Self-Directed Learning, and Demonstrates Teamwork. Soldiers rate the effectiveness of each action on a 7-point scale. Their ratings are compared with the mean ratings of SMEs. These SMEs were 24 senior NCOs attending the U.S. Army Sergeants Major Academy (USASMA).

Scoring the CSJT

Because situational judgment test items are notoriously heterogeneous, we did not compute dimension scores.[9] Rather, we computed an overall CSJT score based on scores assigned to each of the items. A score for each item was computed by taking the mean of the item's four option scores. The item scores were used to compute coefficient alpha.

We computed the judgment score for each response option using the following equation:

$$Judgment\ Score_{Option\ x} = 6 - |\ SoldiersRating_{Option\ x} - keyedEffectiveness_{Option\ x}\ |$$

In the equation above, the keyed effectiveness score for an option was based on the ratings of SMEs. To ensure that Soldiers who gave mid-point ratings to all options were not given an advantage, we adjusted the SME means by "stretching" the range of values and rounding to the nearest integer. (The stretching process is described more fully in Chapter 7, Predictor Situational Judgment Test.) The amount of change depended on the rating's distance from 4. If the mean SME rating was exactly 4, the scale midpoint, then no stretching was done. The farther the rating's distance from 4, the more the rating was changed. (See Waugh & Russell, 2005, for a detailed discussion of the CSJT scoring process). Using the final scoring

[9] Attempts to determine the CSJT's underlying constructs failed. The eigenvalues of the correlation matrix of the option scores suggest between 13 and 18 common factors.

key, a total score of 6.0 is perfect, and a score of .98 is the lowest possible score. On average, a person responding randomly would achieve a score of 3.6, based on simulated random data. Interrater agreement among the 24 SMEs was .976 (.630 for a single SME); interrater reliability was .978 (.650 for a single SME).

Personnel File Form (PFF)

The Personnel File Form (PFF) is a self-report measure that closely parallels the content of the Army NCO Promotion Point Worksheet (PPW) and Personnel File Forms used in past research (e.g., NCO21, Project A). The PPW serves as the basis for the Army's current NCO promotion system. Soldiers receive promotion points in six areas: (a) Commander's Evaluation; (b) Promotion Board Points; (c) Awards, Certificates, and Military Achievements; (d) Military Education; (e) Civilian Education; and (f) Military Training. Promotion points for the first two areas are awarded by a Soldier's commander and promotion board members at the time a Soldier is up for promotion, whereas points for the latter four areas are allocated by the personnel system based on Soldiers' records.

The PFF contains sections that assess Soldiers' standing in the latter four areas of the PPW (i.e., Awards, Certificates, and Military Achievements; Military Education; Civilian Education; and Military Training). Initial content for these sections was drawn from the NCO21 PFF21 (see Knapp, Burnfield et al., 2002). The PFF also asked Soldiers to indicate the number of disciplinary actions (e.g., Article 15s, Flag Actions, arrests) they have been subject to, which should be particularly useful data as criteria for the temperament and P-E fit predictors. In prior research, it was found that Soldiers actually self-reported more negative actions than revealed by their permanent Army records (Riegelhaupt, Harris, & Sadacca, 1987).

Scoring the PFF

We attempted to create scales corresponding to each content area on the PFF. Several of these scales reflected content and scoring algorithms used in past versions of the instrument, while other scales reflected new content for Select21. For new content areas, rational scoring algorithms were developed (see Putka & Campbell, 2005 for scoring details). In total, five PFF scores were analyzed as part of the concurrent validation effort: Awards, Military Education Army Physical Fitness Test, Weapons Qualification, and Disciplinary Actions.[10]

[10] Some of the PFF content described in the measure development report was not used in the concurrent validation effort (Putka & Campbell, 2005; e.g., IET–Exceptional Soldier Designation, Accelerated Advancement to E2). For the most part, the excluded content reflected dichotomous single-item "scales" with unfavorable distributional properties. Given these concerns, and the number of criteria available, we focused our attention on a more limited set of PFF scores that had been examined in past Army research (e.g., NCO21, Project A).

Psychometric Properties of the Performance Criteria

Descriptive Statistics

Both supervisors and peers provided ratings for Soldiers in the concurrent validation. As shown in Table 4.1, 78% of our sample had at least one peer rater, 95% had at least one supervisor rater, and 69% had at least one of both rater types. Only about 7% of Soldiers had two or more supervisor raters; approximately 59% had two or more peer raters. As described in Chapter 2, most supervisor ratings were collected on-site, but some were self-administered after the data collection team left the data collection site. During the field test, we found that inter-rater reliability estimates did not suffer with the inclusion of such "distance" ratings (Keenan, Russell, Le, Katkowski, & Knapp, 2005), which confirmed similar findings in the Army's NCO21 project (Knapp, McCloy, & Heffner, 2004).

Table 4.1. Number of Raters per Soldier

Number of Supervisors	Number of Peers						
	0	1	2	3	4	5	Total
0	44	34	33	34	31	3	179
1	67	78	108	183	130	6	572
2	5	6	16	16	13	1	57
3	0	1	2	0	1	0	4
Total	116	119	159	233	175	10	812

Table 4.2 presents the scale-level descriptive statistics for each criterion measure. The means and *SD*s for both performance rating scales are the average ratings obtained from all available raters (both peers and supervisors) for a Soldier. Scores on the Army-Wide Job Knowledge Test were re-scaled to reflect the percentages of the maximum points.

Reliability Estimates

Measurement error for ratings is assessed by inter-rater reliability which is traditionally estimated by correlating ratings from two different raters for a ratee (cf. Viswesvaran, Ones, & Schmidt, 1996). Because ratees usually have different raters, the conventional approach has been to randomly select and assign the raters into two rater groups, then treat the ratings for all raters in a group as if they came from the same rater for the purpose of estimating reliability. In our analysis we followed the conventional approach, calculating inter-rater reliabilities for peer ratings. However, the assignment of raters into rater groups is often arbitrary, so reliability estimates may vary depending on (a) which rater pair is selected for each rater and (b) how the raters are assigned into the rater groups. Such variation, which reflects the uncertainty of reliability estimates, has generally been ignored in the literature. Therefore, following the conventional approach, we randomly selected and assigned the peer raters into groups to estimate interrater reliabilities. However, we repeated the process 500 times. Table 4.3 presents the results, which are distributions of reliability estimates for peer performance ratings. The means of these distributions provide the best reliability estimates for the ratings, while their *SD*s reflect the variations (i.e., uncertainty) inherent in the traditional approach of arbitrarily assigning raters into rater groups. Reliability

estimates for supervisor ratings were not calculated because of the small number of ratees having two supervisor raters (see Table 4.1). The internal consistency reliability estimates for the other (i.e., non-ratings) criterion measures were .71 for the AWJKT and .91 for the CSJT. Reliabilities for the PFF scales were assumed to be perfect (1.0).

Table 4.2. Criterion Measure Scale-Level Means and SDs

Scale	*n*	*Min*	*Max*	*Mean*	*SD*
Army-Wide Current Observed Performance Ratings [a]					
Common Task Performance	765	1.00	7.00	5.00	0.94
MOS-Specific Task Performance	763	1.00	7.00	4.99	1.06
Communication Performance	768	1.00	7.00	4.79	1.03
Information Management Performance	767	1.00	7.00	4.73	0.99
Problem Solving and Decision Making Performance	767	1.40	7.00	4.64	1.05
Adaptation to Changes	764	1.60	7.00	4.94	0.99
Exhibits Effort and Initiative on the Job	767	1.00	7.00	4.81	1.12
Demonstrates Professionalism and Personal Discipline	767	1.00	7.00	4.80	1.14
Support Peers	768	1.00	7.00	5.13	1.00
Exhibits Tolerance	767	2.00	7.00	5.47	0.89
Demonstrates Personal and Professional Development	768	1.00	7.00	4.76	1.03
Demonstrates Physical Fitness	767	1.00	7.00	4.74	1.28
Overall Effectiveness	765	1.60	7.00	5.06	0.90
Army-Wide Future Expected Performance Ratings [a]					
Individual Pace and Intensity	768	1.40	7.00	4.81	0.96
Learning Environment	767	1.00	7.00	4.96	0.93
Disciplined Initiative	767	1.20	7.00	4.82	1.07
Communication Method and Frequency	768	1.40	7.00	5.02	0.94
Army-Wide Job Knowledge Test	763	27.01	96.61	59.82	11.19
Personnel File Form					
Awards	778	0.00	225.00	37.72	35.15
Military Education	778	0.00	98.00	5.21	11.03
Army Physical Fitness Test	778	14.00	300.00	243.35	35.75
Weapons Qualification	778	0.00	50.00	30.06	15.72
Disciplinary Actions	778	0.00	1.00	0.20	0.28
Criterion Situational Judgment Test	596	3.06	5.37	4.44	0.47

[a] Descriptive statistics for these ratings are the average obtained from all available raters (both peers and Supervisors) for a Soldier.

Table 4.3. Distributions of Inter-Rater Reliabilities for Peer Ratings across 500 Random Data Sets

Performance Rating Scale	Reliability Distribution			
	Min	*Max*	*Mean*	*SD*
Army-Wide Current Observed Performance Ratings				
Common Task Performance	.14	.32	.24	.03
MOS-Specific Task Performance	.14	.30	.22	.03
Communication Performance	.12	.29	.21	.03
Information Management Performance	.14	.31	.22	.03
Problem Solving and Decision Making Performance	.13	.29	.19	.03
Adaptation to Changes	.09	.30	.18	.03
Exhibits Effort and Initiative on the Job	.10	.29	.19	.03
Demonstrates Professionalism and Personal Discipline	.18	.35	.27	.03
Support Peers	.06	.24	.15	.03
Exhibits Tolerance	-.01	.17	.08	.03
Demonstrates Personal and Professional Development	.19	.36	.27	.03
Demonstrates Physical Fitness	.27	.43	.35	.03
Overall Effectiveness	.16	.35	.26	.03
Army-Wide Future Expected Performance Ratings				
Individual Pace and Intensity	.16	.34	.25	.03
Learning Environment	.08	.26	.16	.03
Disciplined Initiative	.11	.27	.19	.03
Communication Method and Frequency	.10	.30	.19	.03

Note. Sample size within each dataset was 569.

Scale Intercorrelations

Although we examined a large number of criterion measures, theoretically we would expect that there are only a few performance factors underlying the scales. In other words, the performance factors should account for the pattern of relationships among the scales. Table 4.4 presents the raw (observed) intercorrelations. These correlations, however, not only reflect the underlying performance factors but are also affected by "method effects." Specifically, as shown in Table 4.4, correlations among performance ratings are indiscriminately high because they are inflated by "halo" effect and correlated measurement error due to common raters. Thus, it is important to control for these method effects to examine the underlying factor structures of the criterion scales. We describe the approach adopted to address the issue in Chapter 5, which also describes our search for the underlying performance factors.

Subgroup Differences

We examined the data for gender differences (see Table 4.5), although the large difference in sample sizes should be noted. Female Soldiers had significantly higher average ratings than males on seven AW COPRS scales, three AW FX scales, and the CSJT. Males scored higher on average on the AWJKT and PFF Weapons Qualification, both of which primarily have content commonly associated with combat MOS.

Table 4.4. Criterion Measure Scale-Level Intercorrelations

		1	2	3	4	5	6	7	8	9	10	11	12	13
Army-Wide Current Observed Performance Ratings														
1	Common Task Performance													
2	MOS-Specific Task Performance	**.71**												
3	Communication Performance	**.60**	**.54**											
4	Information Management Performance	**.61**	**.54**	**.61**										
5	Problem Solving & Decision Making Performance	**.62**	**.57**	**.59**	**.65**									
6	Adaptation to Changes	**.58**	**.51**	**.45**	**.53**	**.61**								
7	Exhibits Effort and Initiative on the Job	**.61**	**.61**	**.47**	**.51**	**.59**	**.59**							
8	Demonstrates Professionalism &Personal Discipline	**.53**	**.45**	**.44**	**.47**	**.54**	**.53**	**.64**						
9	Support Peers	**.43**	**.40**	**.38**	**.38**	**.41**	**.45**	**.50**	**.55**					
10	Exhibits Tolerance	**.33**	**.32**	**.31**	**.33**	**.33**	**.31**	**.33**	**.39**	**.49**				
11	Demonstrates Personal & Professional Development	**.63**	**.58**	**.49**	**.56**	**.58**	**.53**	**.63**	**.64**	**.45**	**.35**			
12	Demonstrates Physical Fitness	**.40**	**.34**	**.26**	**.22**	**.28**	**.30**	**.36**	**.36**	**.26**	**.17**	**.41**		
13	Overall Effectiveness	**.72**	**.67**	**.60**	**.61**	**.68**	**.63**	**.73**	**.68**	**.56**	**.42**	**.72**	**.47**	
Army-Wide Future Expected Performance Ratings														
14	Individual Pace and Intensity	**.62**	**.54**	**.49**	**.55**	**.60**	**.53**	**.58**	**.58**	**.41**	**.30**	**.65**	**.46**	**.72**
15	Learning Environment	**.58**	**.55**	**.55**	**.54**	**.57**	**.47**	**.53**	**.54**	**.39**	**.32**	**.63**	**.33**	**.68**
16	Disciplined Initiative	**.62**	**.56**	**.54**	**.58**	**.66**	**.53**	**.65**	**.64**	**.42**	**.35**	**.69**	**.34**	**.73**
17	Communication Method and Frequency	**.57**	**.53**	**.58**	**.57**	**.65**	**.52**	**.57**	**.54**	**.43**	**.33**	**.62**	**.31**	**.69**
18	Army-Wide Job Knowledge Test	**.18**	**.12**	**.14**	**.13**	**.16**	**.15**	**.14**	**.09**	.07	.02	**.12**	.00	.07
Personnel File Form														
19	Awards	**.13**	**.11**	.02	.07	**.09**	.06	.06	.00	.00	.03	.05	.00	.04
20	Military Education	**.19**	**.14**	**.08**	**.12**	**.12**	**.10**	**.09**	**.09**	.06	**.07**	**.16**	.07	**.12**
21	Army Physical Fitness Test	**.10**	.05	.07	**.07**	**.09**	.06	.05	.04	-.04	-.06	**.10**	**.43**	**.09**
22	Weapon Qualifications	**.11**	**.12**	**.07**	**.10**	.07	**.08**	.04	-.02	**-.07**	-.06	.04	.02	.06
23	Disciplinary Actions	**-.16**	**-.11**	**-.15**	**-.11**	**-.14**	**-.13**	**-.16**	**-.26**	**-.13**	**-.08**	**-.25**	**-.15**	**-.22**
24	Criterion Situational Judgment Test	**.17**	**.12**	**.15**	**.15**	**.13**	**.10**	**.15**	**.16**	.07	.07	**.15**	.07	**.13**

Table 4.4. (Continued)

		14	15	16	17	18	19	20	21	22	23
Army-Wide Future Expected Performance Ratings											
14	Individual Pace and Intensity										
15	Learning Environment	**.70**									
16	Disciplined Initiative	**.75**	**.69**								
17	Communication Method and Frequency	**.70**	**.72**	**.74**							
18	Army-Wide Job Knowledge Test	**.12**	**.15**	**.11**	**.13**						
Personnel File Form											
19	Awards	.03	-.02	.05	.02	.06					
20	Military Education	**.14**	**.12**	**.11**	**.13**	**.07**	**.24**				
21	Army Physical Fitness Test	**.14**	.05	**.10**	.07	.06	**.09**	.02			
22	Weapon Qualifications	.06	.07	**.08**	.05	**.19**	**.20**	.04	**.16**		
23	Disciplinary Actions	**-.22**	**-.15**	**-.24**	**-.19**	.00	.01	.01	**-.13**	.01	
24	Criterion Situational Judgment Test	**.11**	**.15**	**.14**	**.11**	**.18**	.00	**.09**	.04	**-.09**	-.06

Note. n = 562-768. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table 4.5. Criterion Measure Scale-Level Scores by Gender

Scale		Male		Female	
	d_{FM}	*M*	*SD*	*M*	*SD*
Army-Wide Current Observed Performance Ratings					
Common Task Performance	0.08	4.99	0.95	5.07	0.95
MOS-Specific Task Performance	0.16	4.97	1.05	5.14	1.14
Communication Performance	**0.47**	4.75	1.01	5.22	1.11
Information Management Performance	**0.35**	4.69	0.99	5.04	0.96
Problem Solving and Decision Making Performance	**0.28**	4.61	1.03	4.90	1.13
Adaptation to Changes	-0.07	4.94	0.97	4.88	1.15
Exhibits Effort and Initiative on the Job	**0.24**	4.78	1.12	5.05	1.03
Demonstrates Professionalism and Personal Discipline	**0.30**	4.77	1.12	5.10	1.30
Support Peers	0.15	5.11	0.98	5.26	1.17
Exhibits Tolerance	**0.49**	5.43	0.90	5.87	0.73
Demonstrates Personal and Professional Development	**0.42**	4.72	1.03	5.15	0.96
Demonstrates Physical Fitness	-0.18	4.76	1.27	4.54	1.35
Overall Effectiveness	**0.36**	5.02	0.90	5.35	0.87
Army-Wide Future Expected Performance Ratings					
Individual Pace and Intensity	0.07	4.80	0.96	4.87	0.95
Learning Environment	**0.29**	4.93	0.92	5.20	0.98
Disciplined Initiative	**0.36**	4.78	1.06	5.16	1.03
Communication Method and Frequency	**0.37**	4.99	0.92	5.33	1.07
Army-Wide Job Knowledge Test	**-0.57**	60.45	11.15	54.11	9.94
Personnel File Form					
Awards	**-0.27**	38.75	36.00	28.91	24.48
Military Education	0.14	5.07	10.83	6.57	12.78
Army Physical Fitness Test	-0.17	243.89	35.29	238.06	39.53
Weapons Qualification	**-0.67**	31.10	15.70	20.65	12.70
Disciplinary Actions	-0.14	0.21	0.29	0.17	0.25
Criterion Situational Judgment Test	**0.44**	4.42	0.47	4.63	0.39

Note. n_{Female} = 73 - 77. n_{Male} = 685-700. d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

We also looked for subgroup differences due to race/ethnicity, which are shown in Table 4.6. White Soldiers had higher average ratings than Black Soldiers on four AW COPRS scales, three FX scales, the AWJKT, and PFF Weapons Qualification. Black Soldiers had higher average ratings than Whites for the PFF Disciplinary Actions scale. Hispanics had higher average ratings than Whites on three of the AW COPRS scales; Whites had higher average scores than Hispanics on the AWJKT and PFF Weapons Qualification. The subgroup differences on the AWJKT for both Black (d_{BW} = -.81) and Hispanic (d_{HW} = -.51) Soldiers are quite a bit higher than for any of the other scales or instruments. The absolute values of the other significant differences range from 0.19 (FX Learning Environment) to .39 (Weapons Qualification).

Table 4.6. Criterion Measure Scale-Level Scores by Race/Ethnic Group

Scale			White		Black		White Non-Hispanic		Hispanic	
	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Army-Wide Current Observed Performance Ratings										
Common Task Performance	-0.16	0.07	5.02	0.96	4.86	0.93	5.00	0.98	5.08	0.84
MOS-Specific Task Performance	-0.04	0.09	5.00	1.08	4.95	1.05	4.98	1.09	5.07	1.01
Communication Performance	**-0.20**	-0.09	4.84	1.03	4.64	0.95	4.86	1.04	4.77	1.00
Information Management Performance	**-0.21**	0.15	4.77	0.99	4.56	0.97	4.73	1.00	4.88	0.95
Problem Solving and Decision Making Performance	-0.16	0.02	4.67	1.07	4.50	0.99	4.67	1.08	4.69	1.01
Adaptation to Changes	**-0.24**	0.12	4.98	1.00	4.75	0.93	4.96	1.02	5.08	0.90
Exhibits Effort and Initiative on the Job	**-0.22**	-0.04	4.85	1.15	4.60	0.99	4.87	1.18	4.82	1.04
Demonstrates Professionalism & Personal Discipline	-0.17	**0.20**	4.82	1.13	4.62	1.18	4.78	1.17	5.02	1.01
Support Peers	-0.07	0.02	5.14	0.99	5.07	1.06	5.14	1.01	5.16	0.94
Exhibits Tolerance	0.12	**0.28**	5.42	0.88	5.52	0.93	5.38	0.88	5.62	0.84
Demonstrates Personal and Professional Development	-0.16	**0.23**	4.77	1.04	4.61	0.99	4.72	1.07	4.97	0.87
Demonstrates Physical Fitness	0.03	0.13	4.71	1.28	4.75	1.30	4.68	1.26	4.84	1.34
Overall Effectiveness	-0.09	0.16	5.06	0.93	4.98	0.84	5.03	0.96	5.19	0.79
Army-Wide Future Expected Performance Ratings										
Individual Pace and Intensity	-0.08	0.15	4.80	1.01	4.72	0.83	4.79	1.03	4.94	0.89
Learning Environment	**-0.19**	-0.02	5.00	0.95	4.82	0.89	5.00	0.98	4.99	0.83
Disciplined Initiative	**-0.22**	0.14	4.85	1.07	4.62	1.04	4.82	1.10	4.97	0.95
Communication Method and Frequency	**-0.25**	0.01	5.07	0.96	4.83	0.85	5.07	0.98	5.08	0.88
Army-Wide Job Knowledge Test	**-0.86**	**-0.51**	61.92	10.62	52.74	10.07	63.01	10.57	57.67	10.22
Personnel File Form										
Awards	-0.01	0.14	36.81	33.20	36.59	40.42	36.00	33.47	40.76	32.63
Military Education	0.13	0.05	4.85	10.37	6.17	11.99	4.83	10.31	5.39	12.51
Army Physical Fitness Test	-0.04	0.04	243.5	36.22	242.1	36.09	243.4	36.06	245.0	35.89
Weapons Qualification	**-0.39**	**-0.30**	31.44	15.64	25.33	14.32	32.38	15.49	27.70	15.80
Disciplinary Actions	**0.28**	-0.17	0.19	0.27	0.26	0.31	0.19	0.28	0.15	0.25
Criterion Situational Judgment Test	-0.20	-0.12	4.46	0.46	4.37	0.49	4.48	0.46	4.42	0.48

Note. n_{White} = 427 - 550. n_{Black} = 117-150. $n_{Non\text{-}HispanicWhite}$ = 336 - 432. $n_{Hispani}c$ = 111-152.d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-Non-Hispanic White mean difference. Effect sizes calculated as (mean of non-referent group – mean of referent group)/*SD* of referent group. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Summary

The criterion measure scale-level scores generally displayed satisfactory psychometric properties. There was a good amount of score variability. The AWJKT and CSJT showed strong internal-consistency reliabilities. For the most part, with the exception of the AWJKT, the subgroup differences were moderate. Although we attempted to minimize the reading requirements by using graphics and a variety of item formats, these efforts did not eliminate the subgroup differences commonly found on knowledge-based tests (Sackett, Schmitt, Ellingson, & Kabin, 2001).

The intercorrelations for the AW COPRS and FX were inflated by halo effects, as was expected. There were not sufficient supervisor raters to allow us to calculate reliabilities for them. The peer ratings showed higher estimated reliabilities for those dimensions that were more easily observed than those that must be largely inferred (e.g., Demonstrates Physical Fitness is more visible to others than is Adaptation to Changes).

We will describe how we developed criterion composites in Chapter 5. Specifically, we will take a more detailed look at the criterion interrelationships via confirmatory factor analysis, with the intent of identifying a reduced set of performance composites for use in subsequent validation analyses.

CHAPTER 5: PERFORMANCE COMPOSITE SCORES AND RELATIONS AMONG CRITERIA

Huy Le and Dan J. Putka
HumRRO

Overview

Previous chapters summarized the psychometric properties of the performance and attitudinal criterion measures developed for Select21. In this chapter, we describe the steps taken to reduce the performance criteria to a more parsimonious set of composites for use in validating the Select21 predictors. Such a reduction is warranted not only for practical reasons, but also for theoretical reasons, as past research has indicated that the performance domain comprises roughly two to eight latent factors (e.g., Borman & Motowidlo, 1993; Campbell & Knapp, 2001; Campbell, McCloy, Oppler, & Sager, 1993). Thus, the first part of this chapter focuses on modeling the Select21 performance domain to determine whether a theoretically meaningful latent structure underlies the performance criteria discussed in Chapter 4. The second part of this chapter describes the psychometric properties of the performance composites that resulted from this modeling effort. The final part of this chapter focuses on the relationships among the performance composites and attitudinal criteria.

Modeling the Select21 Performance Domain

The approach we took to modeling the Select21 performance domain can be divided into three phases. In the first phase, we modeled the latent structure of the Army-Wide Current Observed Performance Rating Scales (AW COPRS) dimensions. We used these AW COPRS-only models to determine how to treat supervisor and peer ratings for subsequent modeling purposes. Specifically, our goal in this first phase was to determine whether peer and supervisor ratings were interchangeable. Assessing the interchangeability of these types of raters was essential given the implications it had for how we constructed subsequent cross-instrument performance models and estimated the reliability of performance composites (discussed later). In the second phase of modeling, we examined cross-instrument performance models consisting of all "current" performance criteria. Our focus here was on modeling the latent structure of the current performance domain in general (not just ratings). Lastly, in the third phase of modeling, we incorporated data from the Army-Wide Future Expected Performance Rating Scales (AW FX) into the final cross-instrument current performance model identified in the previous phase. The purpose of adding AW FX data to the model was to assess the relationship between current and future performance criteria accounting for methodological artifacts such as criterion unreliability (which attenuates current-future criteria relations) and correlated error arising from common raters (which inflates current-future criteria relations).

Phase 1 Modeling: AW COPRS-Only Models

As a first step in modeling the latent structure underlying the AW COPRS dimensions, we identified several theoretically plausible models from the military and civilian research literatures that could underlie the ratings (e.g., Borman & Motowidlo, 1993; Campbell & Knapp,

2001; Campbell et al., 1993).[11] For the most part, the models we identified were hierarchally nested, differing only in the number of performance factors they specified. These competing models were then tested to identify the model that best explained the latent structure underlying the AW COPRS dimensions. When evaluating the relative performance of these models, we considered general model fit (e.g., as indexed by CFI, RMSEA, SRMR, and NNFI) and the reasonableness of model parameter estimates.

In addition to including factors corresponding to latent performance constructs in these models, we also included factors corresponding to rater factors. As discussed in Chapter 4, the nature of the Select21 ratings measurement design was such that raters provided ratings on all AW COPRS dimensions for a given Soldier (what authors in the Generalizability theory literature have referred to as a *linked* measurement design; Brennan, 2001). As such, the covariation between any two dimensions may reflect true covariance attributable to some higher level performance construct, but also correlated error arising from having the same rater provide ratings on both dimensions (e.g., Scullen, Mount, & Goff, 2000). Failing to take the correlated errors into account when modeling the latent structure of the performance ratings arising from a linked measurement design would weaken one's ability to find meaningful latent performance factors. Therefore, we modeled the performance ratings at the disaggregate level (e.g., two variables for each AW COPRS dimension were included in the analysis, one for "Rater 1" and another for "Rater 2"). Basing models on disaggregated data allowed us to include latent factors representing rater effects in the model. Appendix A provides an example of a disaggregated correlation matrix.

In fitting the aforementioned models to the AW COPRS data, a key decision was how to assign raters as "Rater 1" and "Rater 2" to each Soldier for modeling purposes. Given that each Soldier was potentially rated by up to five peers and three supervisors and that the raters for each Soldier were not necessarily the same across Soldiers, this decision was not straightforward. Moreover, the ill-structured nature of this measurement design gave rise to two other issues. The first issue was how best to assess the interchangeability of peer and supervisor raters. The second issue was how best to account for the arbitrariness of the solutions we would get by following the standard practice in the literature of randomly assigning raters to be "Rater 1" and "Rater 2" (e.g., Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen et al., 2000).

To resolve the first issue, we fitted models in which "Rater 1" was required to be a peer rater and in which "Rater 2" was required to be a supervisor rater. Within these models we specified a Peer factor, on which all AW COPRS dimensions rated by "Rater 1" (the peer rater) loaded, and a Supervisor factor, on which all AW COPRS dimensions rated by "Rater 2" (the supervisor rater) loaded. Next, we compared the fit of two competing versions of each performance model. In the first, we constrained all factor loadings for Peer to be equal to the corresponding loadings for Supervisor. In the second, we allowed these loadings to be freely estimated for the Peer and Supervisor factors. Comparing these two models enabled us to test the hypothesis that peer and supervisor raters were interchangeable.

[11] The AW COPRS Overall Performance scale was omitted from all modeling analyses discussed in this chapter. Unlike the other AW COPRS scales, which focused on specific dimensions of Army-wide job performance (e.g., Supporting Peers), the Overall Performance scale focused on the Soldier's performance in general. Given its breadth of focus, we did not feel that it made conceptual sense to include it in models designed to examine the latent structure of the performance domain (assuming that domain comprised more than one factor).

Our resolution of the second issue was to use a resampling strategy similar to the one described in Chapter 4 for estimating interrater reliabilities for single AW COPRS and AW FX performance dimensions. Like the previous chapter, had we simply randomly chosen a peer rater for each Soldier to serve as "Rater 1" and a supervisor rater to serve as "Rater 2" for purposes of fitting the confirmatory factor analysis (CFA) models, the observed correlation between any pair of rating dimension variables would have been completely arbitrary (just as the single-rater reliability estimates based on any single sampling of the data were in the previous chapter). Indeed, the arbitrary results obtained from randomly selecting a pair of raters for each ratees would be even more widespread in our analyses because CFAs are conducted on matrices of correlations (or covariances). Therefore, more than one correlation (e.g., an estimate of single rater reliability) would be affected by the random assignment of raters.[12] To address this problem, we created 500 modeling samples by randomly selecting and assigning raters to the *same sample of ratees*, and then carried out analyses (i.e., fitting CFA models) within each sample. We then aggregated statistics (i.e., fit statistics, standardized loadings) across these 500 samples to draw conclusions about the appropriateness of the performance models. Arguably, this approach reduces the uncertainty resulting from the arbitrariness of rater assignment and selection process typically used in the literature (e.g., Mount et al., 1998; Scullen et al., 2000).

In sum, we formulated five competing models with different numbers of performance factors underlying the AW COPRS ratings. We examined a model with one general performance factor, a model with two factors (Can-Do vs. Will-Do performance), two models with three-factors, and a four-factor model. As noted previously, most of these models were hierarchically nested (except for the two alternative three-factor models), so their relative fit could be examined using chi-squared difference tests (Widaman, 1985). For each of the five performance rating models, we specified two hierarchically nested sub-models which were different in how ratings from different sources were treated. The first sub-model allowed ratings from supervisor and peer to be different (i.e., freely estimated); the second sub-model constrained the ratings (for each dimension) to be the same across the two sources. Altogether, 10 partially hierarchically nested models were tested. In all models we fitted, the covariances among the performance factors were free to vary, but the covariances between the rater factors and performance factors were constrained to zero, and the covariance between the rater factors was constrained to zero (viewing the rater factors as representing sources of idiosyncratic, rater-specific variance).

[12] As illustrated in results presented later, the arbitrariness of results obtained by following standard practices in the literature for dealing with such data (i.e., basing the CFA on a single random selection and assignment of raters to "Rater 1" and "Rater 2" variables for each ratee) is substantial. This arbitrariness is evidenced by the wide range of standardized factor loadings observed across samples based on different selection and assignment of raters (even though those samples were based on the same exact sample of ratees). This variation cannot be explained by traditional notions of sampling error because the same group of ratees was analyzed, and is not wholly accounted for by the sampling of raters either, as part of the effect stems from how the raters chosen were assigned to "Rater 1" and "Rater 2" columns (not simply just which raters were chosen for each ratee).

Phase 1 Modeling Results

When fitting the models, we conducted analyses using the Wave 1 sample only. Though all the models we examined yielded reasonable levels of fit, the best model appeared to be the four-factor model that allowed parameters for peer and supervisor factors to vary freely. This model provided a good fit to the data (average RMSEA = .048; average CFI = .953). Figure 5.1 shows the path diagram for the final AW COPRS model, and Tables 5.1 and 5.2 show final model fit statistics and standardized loadings, respectively. Examination of loadings for peer and supervisor factors revealed notable differences for AW COPRS dimensions underlying the Effort and Initiative (EI) and Teamwork (TEAM) factors.[13]

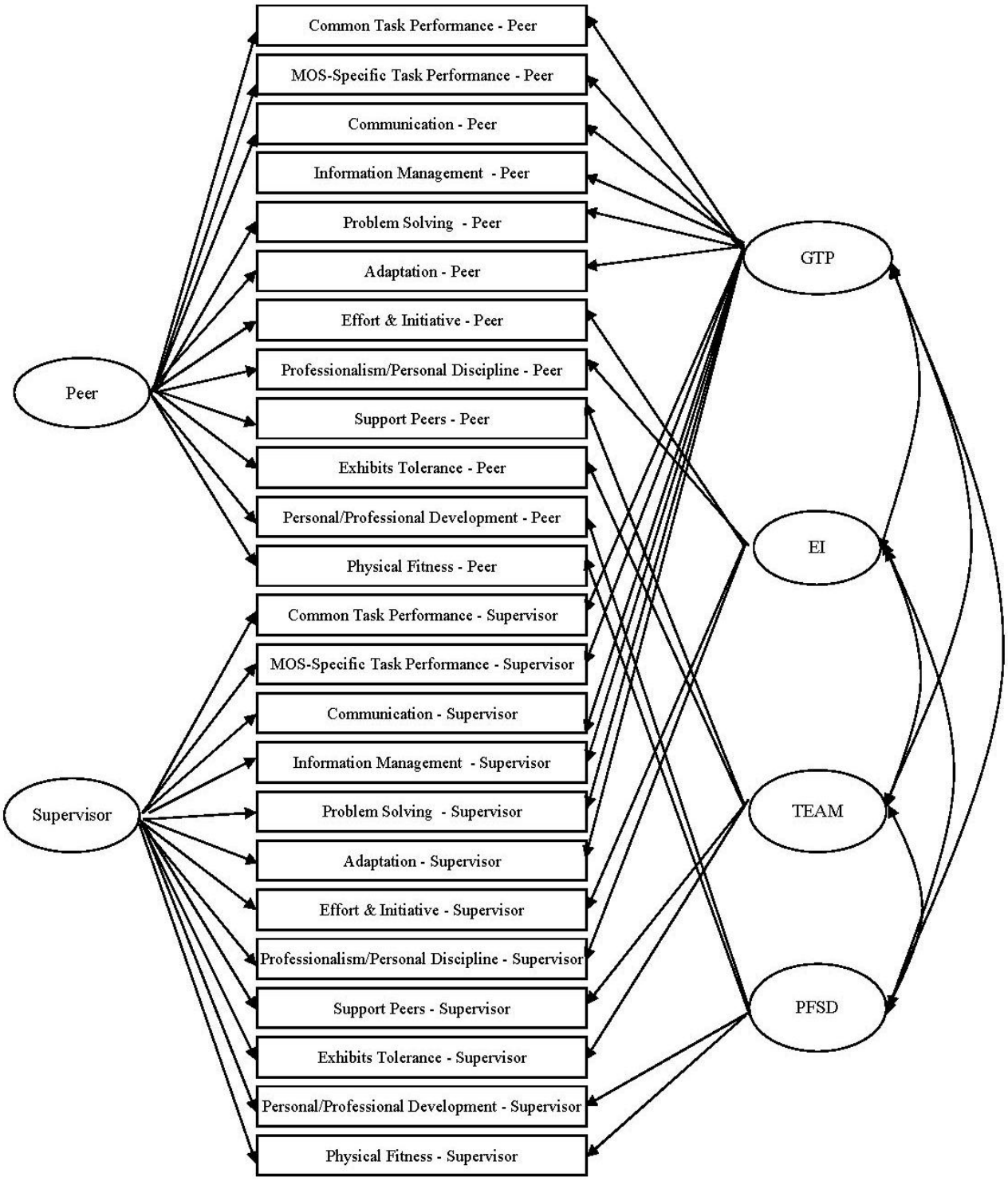
Phase 2 Modeling: Cross-Instrument Current Performance Models

We formulated cross-instrument performance models by adding non-rating measures to the final ratings-only model, and specifying latent factors underlying those non-rating measures that were most theoretically appropriate (e.g., the General Technical Proficiency factor was specified to underlie the Army-Wide Job Knowledge Test [AWJKT]). The addition of non-rating measures also allowed us to specify several alternative models which were not possible for models with only ratings (e.g., creating a factor that focused solely on physical fitness and comprised AW COPRS Physical Fitness and Army Physical Fitness Test [APFT] scores from the Personnel File Form [PFF]). As with the AW COPRS-only models, we compared the relative fit of several different competing models. Table 5.3 shows the range of performance models we considered.[14]

The choice and range of these models was largely influenced by performance models examined as part of Project A (Campbell, Hanson, & Oppler, 2001). Additionally, the models reflect varying degrees of specificity in terms of the level at which the performance constructs are operationalized. For example, at a very general level, we hypothesized that a two-factor model might underlie the data, reflecting "can do" and "will do" performance criteria (Campbell et al., 2001). We also split the two-factor model into a three-factor model based on distinctions among different types of will-do performance. This model specified General Technical Proficiency, Achievement and Effort, and Physical Fitness and Self Development as factors. It roughly corresponds to the three-factor model discussed by Campbell et al. (2001) that specified "Can Do" performance, Achievement, Leadership, Personal Discipline, and Physical Fitness/Military Bearing as factors.

[13] Note, we also fit an analogous version of this model based on peer raters only (i.e., the Peer and Supervisor factors in this model were replaced with a "Peer 1" and a "Peer 2" factor). This peer-only model revealed similar patterns of loadings for Peer 1 and Peer 2 (constraining the peer loadings to equality did not result in substantially poorer model fit). Thus, the differences observed between the Peer and Supervisor loadings in the model described here appear to reflect differences due to rater perspective, rather than just individual rater idiosyncrasies.

[14] Although discussed in Chapter 4, the PFF Awards scale was omitted from the cross-instrument models discussed in this chapter. Initially, we had hypothesized PFF Awards would load on an Achievement and Effort factor (discussed below). However our initial modeling efforts suggested PFF Awards did not load on this factor (nor did it load on any other factor). Therefore, we excluded it from subsequent modeling analyses.



GTP = General Technical Proficiency, EI = Effort and Initiative, TEAM = Teamwork, PFSD = Physical Fitness and Self-Development.

Figure 5.1. Final AW COPRS model.

Table 5.1. Final AW COPRS Model Fit Statistics

Statistic	*M*	*SD*
χ^2	439	38
Degrees of Freedom	222	-
Number of Parameters Estimated	78	-
p-value for χ^2	.000	.000
RMSEA	.048	.004
SRMSR	.066	.019
CFI	.953	.008
NNFI	.942	.010

Note. $n = 424$. *M* = Mean statistic across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of statistic across the 500 samples.

Table 5.2. Final AW COPRS Model Standardized Loadings

	Peer			Supervisor		
Path	*M*	*SD*	SE_M	*M*	*SD*	SE_M
COPRS Common Task Performance- GTP	**.439**	.130	.052	**.357**	.179	.049
COPRS Common Task Performance- RATER	**.610**	.089	.048	**.717**	.121	.044
COPRS Common Task Performance- Residual	**.412**	.047	.036	**.311**	.013	.027
COPRS MOS-Specific Task Performance- GTP	**.463**	.120	.053	**.324**	.146	.052
COPRS MOS-Specific Task Performance- RATER	**.544**	.091	.049	**.628**	.107	.046
COPRS MOS-Specific Task Performance- Residual	**.467**	.051	.040	**.468**	.011	.036
COPRS Communication- GTP	**.389**	.101	.055	**.314**	.130	.055
COPRS Communication- RATER	**.520**	.075	.049	**.536**	.098	.048
COPRS Communication- Residual	**.563**	.044	.044	**.588**	.009	.044
COPRS Info Management- GTP	**.364**	.143	.053	**.377**	.145	.052
COPRS Info Management- RATER	**.569**	.098	.049	**.611**	.105	.046
COPRS Info Management- Residual	**.514**	.041	.041	**.453**	.019	.036
COPRS Problem Solving- GTP	**.366**	.127	.054	**.326**	.167	.052
COPRS Problem Solving- RATER	**.568**	.087	.049	**.640**	.121	.046
COPRS Problem Solving- Residual	**.520**	.039	.041	**.441**	.012	.035
COPRS Adaptation- GTP	**.293**	.099	.055	**.244**	.150	.053
COPRS Adaptation- RATER	**.541**	.063	.049	**.640**	.087	.046
COPRS Adaptation- Residual	**.608**	.033	.045	**.501**	.012	.038
COPRS Effort/Initiative- AE	**.249**	.141	.052	**.342**	.181	.046
COPRS Effort/Initiative- RATER	**.669**	.067	.047	**.709**	.103	.045
COPRS Effort/Initiative- Residual	**.466**	.043	.040	**.337**	.022	.030
COPRS Professionalism/Personal Discipline- AE	**.336**	.150	.054	**.529**	.149	.052
COPRS Professionalism/Personal Discipline- RATER	**.660**	.069	.046	**.683**	.083	.046
COPRS Professionalism/Personal Discipline- Residual	**.424**	.071	.042	**.225**	.123	.036

Table 5.2. (continued)

Path	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
COPRS Supports Peers- TEAM	**.244**	.275	.064	**.591**	.309	.066
COPRS Supports Peers- RATER	**.579**	.093	.049	**.612**	.128	.049
COPRS Supports Peers- Residual	**.521**	.186	.069	**.165**	.159	.075
COPRS Exhibits Tolerance- TEAM	**.135**	.118	.057	**.343**	.185	.056
COPRS Exhibits Tolerance- RATER	**.418**	.049	.051	**.502**	.092	.050
COPRS Exhibits Tolerance- Residual	**.791**	.058	.063	**.588**	.089	.052
COPRS Personal/Professional Development- PFSD	**.302**	.132	.055	**.358**	.211	.052
COPRS Personal/Professional Development- RATER	**.656**	.058	.047	**.688**	.093	.045
COPRS Personal/Professional Development- Residual	**.458**	.034	.040	**.346**	.036	.036
COPRS Physical Fitness- PFSD	**.400**	.147	.071	**.433**	.144	.075
COPRS Physical Fitness- RATER	**.407**	.062	.051	**.338**	.095	.049
COPRS Physical Fitness- Residual	**.649**	.122	.068	**.669**	.138	.076

Note. $n = 424$. *M* = Mean standardized loading across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of standardized loadings across the 500 samples. SE_M = Mean standard error of the standardized loadings across the 500 samples. Bolded loadings were statistically significant ($p < .05$, two tailed, based on average loading-to-*SE* ratio across samples).

Phase 2 Modeling Results

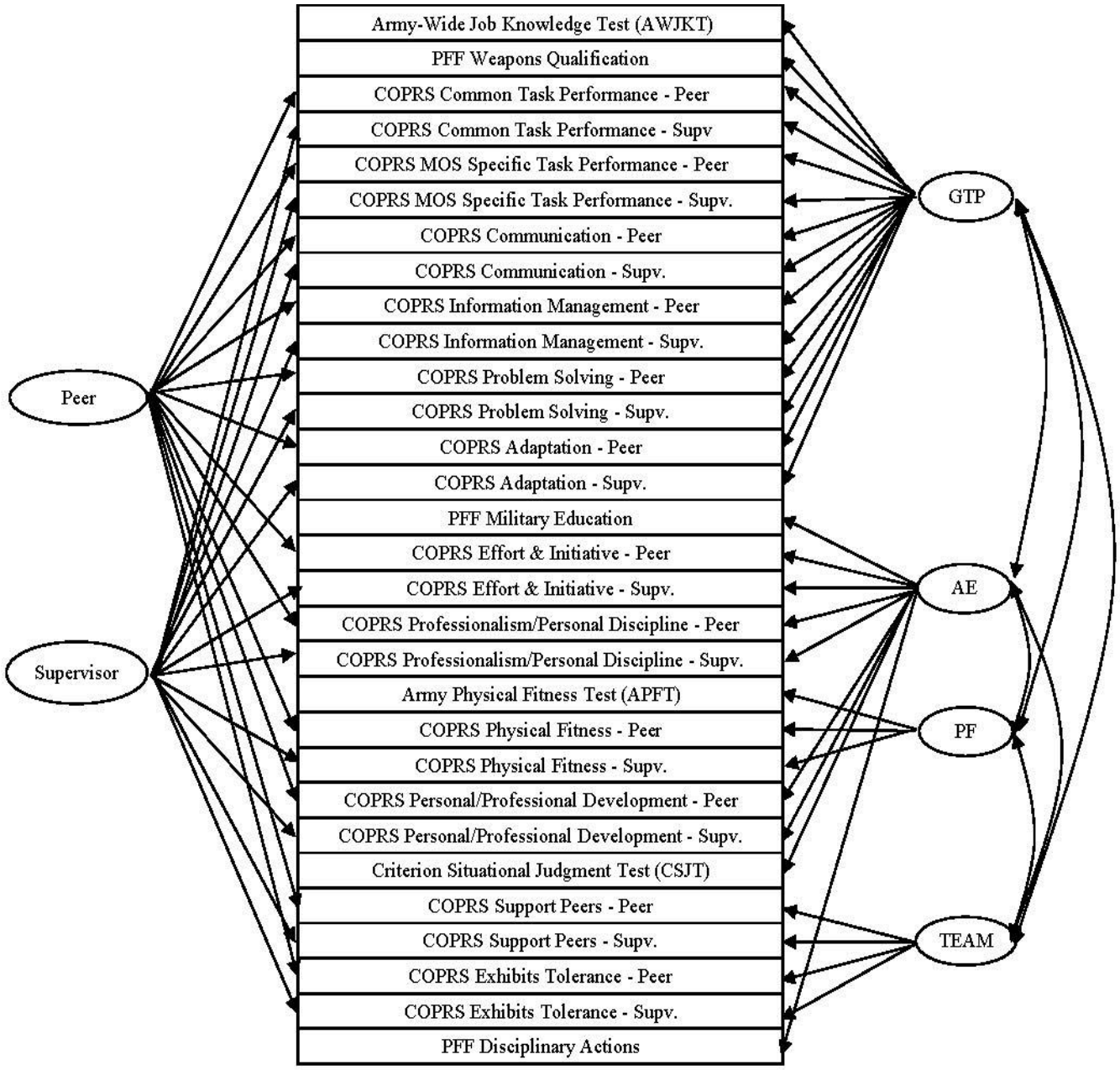
As we did with the AW COPRS-only models, when fitting the cross-instrument models, we first conducted analyses using the Wave 1 sample only. (The mean correlations obtained from this exercise are shown in Appendix A.) Based on these analyses we identified the best model and then used the Wave 2 data to assess the extent to which the model cross-validated. Modeling results indicated that a cross-instrument current performance model with four performance factors (Model 6 in Table 5.3) provided the best fit to the data (average RMSEA = .040; average CFI = .951). Figure 5.2 shows the path diagram for the final model, and Tables 5.4 and 5.5 show final model fit statistics and standardized loadings, respectively. Perhaps the most striking aspect of the results presented in these tables is the low loadings for the non-rating indicators (e.g., Criterion Situational Judgment Test [CSJT], AWJKT). One potential reason for the low loadings of these indicators is the possibility that latent performance factors are saturated with ratings-specific variance. Despite the fact that rater perspective-specific factors were included in the model, a general "ratings method" factor was not. The latent performance factors thus may have been heavily saturated with ratings-specific variance.[15]

[15] To test this possibility, for every model shown in Table 5.3, we examined an additional model with one general rating factor underlying all the ratings (beyond the supervisor and peer rating factors). However, results obtained from these models were very similar to models without that general rating factor in terms of fit indexes and loadings of non-rating factors. Thus, the reason why non-ratings have relatively low loadings in the model remains unclear.

Table 5.3. Cross-Instrument Current Performance Models Tested

	Criterion Performance Models										
	2-Factors	3-Factors		4-Factors			5-Factors				
Performance Measures	1	2	3	4	5	6	7	8	9	10	11
COPRS Common Task Performance	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS MOS-Specific Task Performance	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS Communication Performance	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS Information Management	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS Problem Solving	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS Adaptation to Changes	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
COPRS Effort/Initiative	AE	AE	AE	AE	AE	AE	AE	AE	AE	AE	AE
COPRS Professionalism/Personal Discipline	AE	AE	AE	AE	AE	AE	DYSF	AE	DYSF	DYSF	DYSF
COPRS Support Peers	AE	TEAM	AE	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM
COPRS Exhibits Tolerance	AE	TEAM	AE	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM	TEAM
COPRS Personal/Professional Development	AE	AE	PFSD	AE	PFSD	AE	PFSD	DYSF	AE	AE	AE
COPRS Physical Fitness	AE	AE	PFSD	PF	PFSD	PF	PFSD	PF	PF	PF	PF
Army-Wide Job Knowledge Test (AWJKT)	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
PFF Military Education	AE	AE	AE	AE	AE	AE	AE	AE	AE	AE	AE
PFF Army Physical Fitness Test	AE	AE	PFSD	PF	PFSD	PF	PFSD	PF	PF	PF	PF
PFF Weapons Qualification	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP	GTP
PFF Disciplinary Actions	AE	AE	AE	AE	AE	AE	DYSF	DYSF	DYSF	DYSF	DYSF
PFF Awards	AE	AE	AE	AE	AE	*	AE	AE	AE	GTP	GTP
Criterion Situational Judgment Test (CSJT)	AE	TEAM	AE	AE	TEAM	AE	TEAM	TEAM	GTP	DYSF	AE

Note. GTP = General Technical Proficiency, AE = Achievement and Effort, TEAM = Teamwork, PFSD = Physical Fitness and Self Development, PF = Physical Fitness, DYSF = Dysfunctional Behaviors.



GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork.

Figure 5.2. Final cross-instrument current performance model.

Table 5.4. Final Cross-Instrument Current Performance Model Fit Statistics

Statistic	*M*	*SD*
χ^2	599	32
Degrees of Freedom	375	-
Number of Parameters Estimated	90	-
p-value for χ^2	.000	.000
RMSEA	.040	.003
SRMSR	.060	.007
CFI	.951	.007
NNFI	.943	.008

Note. n = 375. Sample size is smaller than that in the rating-only model because a number of Soldiers do not have the CSJT scores. *M* = Mean statistic across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of statistic across the 500 samples.

Table 5.5. Final Cross-Instrument Current Performance Model Standardized Loadings

Type of Measure/Path	*M*	*SD*	SE_M
Non-Rating Measures			
AWJKT- GTP	**.147**	.042	.065
AWJKT- Residual	**.977**	.013	.072
PFF Weapons Qualification- GTP	.114	.043	.065
PFF Weapons Qualification- Residual	**.985**	.010	.073
CSJT- AE	**.153**	.029	.068
CSJT- Residual	**.976**	.009	.073
APFT- PF	**.553**	.037	.065
APFT- Residual	**.693**	.041	.071
PFF Disciplinary Actions- AE	**-.397**	.043	.068
PFF Disciplinary Actions- Residual	**.841**	.035	.072
PFF Military Education- AE	**.156**	.032	.068
PFF Military Education- Residual	**.974**	.010	.073

	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
Ratings						
COPRS Common Task Performance- GTP	**.525**	.067	.054	**.316**	.070	.055
COPRS Common Task Performance- RATER	**.564**	.055	.052	**.749**	.038	.044
COPRS Common Task Performance- Residual	**.390**	.044	.036	**.312**	.010	.028
COPRS MOS-Specific Task Performance- GTP	**.502**	.069	.056	**.274**	.068	.057
COPRS MOS-Specific Task Performance- RATER	**.534**	.058	.053	**.692**	.036	.047
COPRS MOS-Specific Task Performance- Residual	**.447**	.042	.040	**.424**	.009	.035
COPRS Communication- GTP	**.473**	.062	.058	**.301**	.053	.059
COPRS Communication- RATER	**.440**	.058	.055	**.577**	.030	.049
COPRS Communication- Residual	**.569**	.046	.047	**.557**	.010	.044

Table 5.5. (Continued)

Type of Measure/Path	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
Ratings (Continued)						
COPRS Info Management- GTP	**.448**	.077	.057	**.334**	.058	.056
COPRS Info Management- RATER	**.510**	.060	.054	**.670**	.033	.046
COPRS Info Management- Residual	**.523**	.041	.044	**.414**	.009	.034
COPRS Problem Solving- GTP	**.485**	.073	.057	**.258**	.062	.057
COPRS Problem Solving- RATER	**.516**	.059	.054	**.699**	.037	.047
COPRS Problem Solving- Residual	**.482**	.038	.042	**.424**	.013	.035
COPRS Adaptation- GTP	**.370**	.060	.059	**.246**	.059	.058
COPRS Adaptation- RATER	**.487**	.050	.055	**.680**	.030	.047
COPRS Adaptation- Residual	**.614**	.033	.049	**.457**	.009	.037
COPRS Effort/Initiative- AE	**.314**	.063	.060	**.362**	.057	.054
COPRS Effort/Initiative- RATER	**.672**	.042	.051	**.703**	.030	.045
COPRS Effort/Initiative- Residual	**.437**	.039	.039	**.329**	.009	.029
COPRS Professionalism/Personal Discipline- AE	**.345**	.053	.060	**.421**	.051	.054
COPRS Professionalism/Personal Discipline- RATER	**.641**	.035	.051	**.670**	.029	.045
COPRS Professionalism/Personal Discipline- Residual	**.460**	.032	.040	**.323**	.015	.030
COPRS Supports Peers- TEAM	.073	.118	.046	**.760**	.141	.058
COPRS Supports Peers- RATER	**.677**	.045	.052	**.548**	.055	.049
COPRS Supports Peers- Residual	**.519**	.089	.054	**.058**	.091	.057
COPRS Exhibits Tolerance- TEAM	.085	.052	.052	**.398**	.075	.057
COPRS Exhibits Tolerance- RATER	**.463**	.050	.055	**.425**	.036	.053
COPRS Exhibits Tolerance- Residual	**.772**	.046	.061	**.637**	.023	.053
COPRS Personal/Professional Development- AE	**.404**	.064	.059	**.377**	.058	.055
COPRS Personal/Professional Development- RATER	**.619**	.047	.051	**.666**	.033	.046
COPRS Personal/Professional Development- Residual	**.440**	.033	.039	**.369**	.014	.032
COPRS Physical Fitness- PF	**.547**	.043	.060	**.581**	.045	.061
COPRS Physical Fitness- RATER	**.466**	.043	.050	**.419**	.030	.048
COPRS Physical Fitness- Residual	**.493**	.041	.057	**.482**	.042	.061

Note. $n = 375$. *M* = Mean standardized loading across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of standardized loadings across the 500 samples. SE_M = Mean standard error of the standardized loadings across the 500 samples. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork. Bolded loadings were statistically significant ($p < .05$, two tailed, based on average loading-to-*SE* ratio across samples).

Phase 3 Modeling: Current vs. Future Performance

The next step in the modeling process involved taking the final cross-instrument current performance model and adding data from the four AW FX rating scales to it. We specified only one general future performance factor underlying the AW FX rating scales added to the model. Figure 5.3 shows the path diagram for the final model, and Tables 5.6 and 5.7 show final model fit statistics and standardized loadings, respectively. This model allowed us to estimate the relationships between future and current performance factors (which are presented later). Examination of the aforementioned tables reveals that this model fitted the data extremely well (average RMSEA = .045; average CFI = .935). The pattern of loadings was similar to the final cross-instrument performance model.

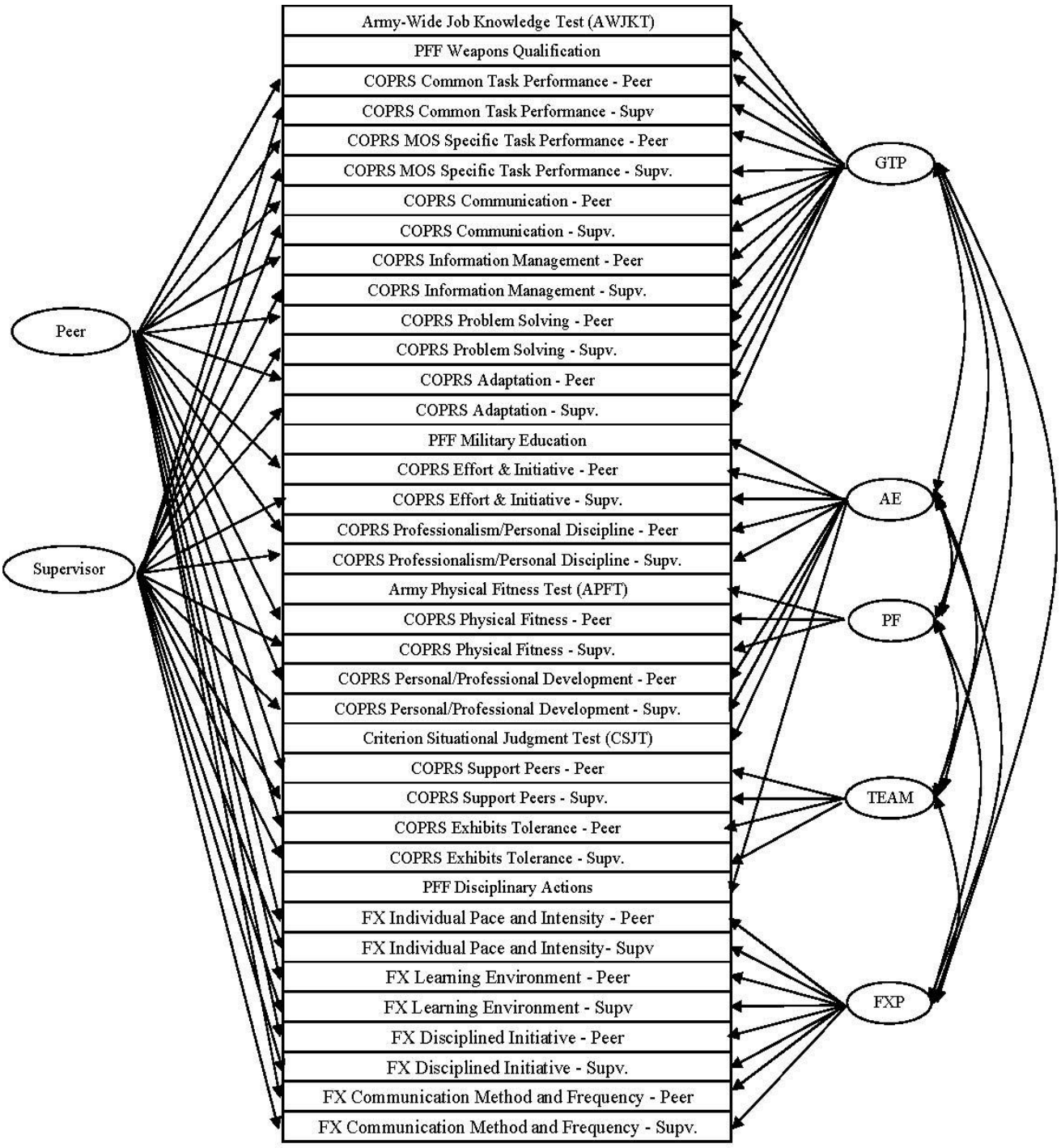
Scoring of Performance Composites

Current and future performance composites were created based on the modeling results. Specifically, we combined all of the criterion measures that loaded on the same underlying performance factor to form a composite representing that factor. The combination process involved two steps: (a) standardizing all of the component performance measures, and (b) averaging the resulting components together to form a composite.

For non-rating measures, the standardization was straightforward. However, for ratings, it was more complicated because (a) most ratees have ratings from supervisors and peers, and (b) the numbers of raters from each source varied across ratees. Because of this measurement design, it was difficult to determine the values of means and standard deviations to be used for standardization. To address this problem, we adopted a solution similar to that used in our modeling approach; that is, we used values obtained from 500 samples created by randomly re-sampling the raters. For each rating dimension, we calculated the means of means and standard deviations of 500 samples of supervisor and peer ratings separately. These values were then used to standardize all the ratings obtained from supervisors and peers. In other words, all the supervisor ratings for a dimension were first subtracted from the mean of means across 500 samples, and then they were divided by the mean of standard deviations for that dimension. A similar procedure was used for peer ratings for that dimension. Next, we averaged all (standardized) supervisor ratings and (standardized) peer ratings for each Soldier in each dimension to obtain the average ratings for each rating source. Finally, these ratings were averaged for each ratee. This final value is the (nominal) standardized rating used for combining (averaging) with non-rating measures.[16]

We created two composite scores for the Achievement and Effort (AE) factor. The first composite (AE1) includes scores on the CSJT, whereas the second (AE2) does not include the CSJT score. The AE2 composite was used to validate the Predictor Situational Judgment Test (PSJT) because the common method component of the AE1 probably would have artificially inflated the observed validity coefficient.

[16] The values resulting from this process have standard deviations smaller than one, so they are not strictly standardized.



GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance.

Figure 5.3. Final cross-instrument current and future performance model path diagram.

Table 5.6. Final Cross-Instrument Current and Future Performance Model Fit Statistics

Statistic	*M*	*SD*
χ^2	1,087	44
Degrees of Freedom	623	-
Number of Parameters Estimated	118	-
p-value for χ^2	.000	.000
RMSEA	.045	.002
SRMSR	.069	.011
CFI	.935	.006
NNFI	.926	.007

Note. n = 370. Sample size is smaller than that in the rating-only model because a number of Soldiers do not have the CSJT scores. *M* = Mean statistic across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of statistic across the 500 samples.

Table 5.7. Final Cross-Instrument Current and Future Performance Model Standardized Loadings

Type of Measure/Path	*M*	*SD*	SE_M
Non-Rating Measures			
AWJKT- GTP	.110	.071	.061
AWJKT- Residual	**.983**	.016	.073
PFF Weapons Qualification- GTP	.084	.056	.061
PFF Weapons Qualification- Residual	**.990**	.011	.073
CSJT- AE	**.156**	.040	.065
CSJT- Residual	**.974**	.013	.073
APFT- PF	**.516**	.045	.065
APFT- Residual	**.732**	.046	.071
PFF Disciplinary Actions- AE	**-.354**	.092	.066
PFF Disciplinary Actions- Residual	**.867**	.070	.072
PFF Military Education- AE	**.154**	.035	.065
PFF Military Education- Residual	**.975**	.010	.073

	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
Ratings						
COPRS Common Task Performance- GTP	**.400**	.098	.053	**.415**	.117	.053
COPRS Common Task Performance- RATER	**.647**	.062	.047	**.678**	.086	.047
COPRS Common Task Performance- Residual	**.408**	.045	.036	**.308**	.009	.027
COPRS MOS-Specific Task Performance- GTP	**.379**	.081	.054	**.371**	.090	.055
COPRS MOS-Specific Task Performance- RATER	**.611**	.056	.048	**.628**	.068	.049
COPRS MOS-Specific Task Performance- Residual	**.474**	.039	.039	**.423**	.011	.034
COPRS Communication- GTP	**.341**	.089	.056	**.335**	.090	.058
COPRS Communication- RATER	**.560**	.055	.050	**.541**	.064	.051
COPRS Communication- Residual	**.559**	.041	.045	**.557**	.009	.044

Table 5.7. (Continued)

Type of Measure/Path	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
Ratings (Continued)						
COPRS Info Management- GTP	**.298**	.071	.056	**.402**	.096	.055
COPRS Info Management- RATER	**.604**	.047	.049	**.618**	.075	.049
COPRS Info Management- Residual	**.540**	.034	.043	**.407**	.010	.033
COPRS Problem Solving- GTP	**.370**	.101	.055	**.353**	.119	.056
COPRS Problem Solving- RATER	**.584**	.064	.049	**.632**	.089	.049
COPRS Problem Solving- Residual	**.507**	.042	.042	**.423**	.014	.035
COPRS Adaptation- GTP	**.295**	.070	.057	**.347**	.120	.056
COPRS Adaptation- RATER	**.539**	.050	.051	**.612**	.088	.049
COPRS Adaptation- Residual	**.615**	.034	.048	**.453**	.009	.037
COPRS Effort/Initiative- AE	**.313**	.080	.055	**.408**	.149	.054
COPRS Effort/Initiative- RATER	**.639**	.054	.048	**.647**	.114	.047
COPRS Effort/Initiative- Residual	**.477**	.040	.040	**.336**	.012	.030
COPRS Professionalism/Personal Discipline- AE	**.350**	.066	.055	**.491**	.134	.053
COPRS Professionalism/Personal Discipline- RATER	**.614**	.043	.048	**.601**	.118	.048
COPRS Professionalism/Personal Discipline- Residual	**.487**	.035	.041	**.317**	.023	.030
COPRS Supports Peers- TEAM	**.115**	.089	.048	**.758**	.091	.056
COPRS Supports Peers- RATER	**.558**	.049	.051	**.495**	.099	.051
COPRS Supports Peers- Residual	**.665**	.044	.053	**.125**	.104	.054
COPRS Exhibits Tolerance- TEAM	**.118**	.072	.054	**.472**	.076	.057
COPRS Exhibits Tolerance- RATER	**.436**	.047	.053	**.373**	.075	.054
COPRS Exhibits Tolerance- Residual	**.788**	.039	.060	**.609**	.034	.054
COPRS Personal/Professional Development- AE	**.324**	.067	.054	**.408**	.111	.054
COPRS Personal/Professional Development- RATER	**.664**	.042	.047	**.639**	.080	.048
COPRS Personal/Professional Development- Residual	**.440**	.031	.037	**.363**	.013	.031
COPRS Physical Fitness- PF	**.527**	.045	.060	**.607**	.059	.063
COPRS Physical Fitness- RATER	**.448**	.042	.048	**.418**	.039	.049
COPRS Physical Fitness- Residual	**.520**	.043	.057	**.447**	.055	.064
FX Individual Pace and Intensity- FXP	**.334**	.142	.051	**.489**	.145	.052
FX Individual Pace and Intensity- RATER	**.672**	.071	.047	**.660**	.111	.048
FX Individual Pace and Intensity- Residual	**.409**	.046	.036	**.264**	.006	.024
FX Learning Environment- FXP	**.287**	.140	.052	**.486**	.183	.052
FX Learning Environment- RATER	**.691**	.068	.047	**.640**	.133	.048
FX Learning Environment- Residual	**.414**	.040	.036	**.276**	.015	.026
FX Disciplined Initiative- FXP	**.338**	.138	.051	**.483**	.170	.052
FX Disciplined Initiative- RATER	**.684**	.070	.047	**.664**	.126	.047
FX Disciplined Initiative- Residual	**.393**	.041	.035	**.252**	.008	.024

Table 5.7. (Continued)

Type of Measure/Path	Peer			Supervisor		
	M	*SD*	SE_M	*M*	*SD*	SE_M
Ratings (Continued)						
FX Communication Method and Frequency- FXP	**.311**	.139	.052	**.500**	.192	.053
FX Communication Method and Frequency- RATER	**.655**	.070	.048	**.612**	.145	.049
FX Communication Method and Frequency- Residual	**.448**	.040	.038	**.290**	.012	.028

Note. $n = 370$. *M* = Mean standardized loading across 500 samples created by randomly sampling one peer rater and one supervisor rater for each Soldier in each sample. *SD* = Standard deviation of standardized loadings across the 500 samples. SE_M = Mean standard error of the estimated standardized loadings across the 500 samples. Bolded loadings were statistically significant ($p < .05$, two tailed, based on average loading-to-*SE* ratio across samples).

Psychometric Properties of Performance Composites

Table 5.8 provides basic descriptive statistics for the performance composites. As mentioned in the previous section, the composites were formed by averaging standardized criterion scores. Therefore, the means of the composites were essentially equal to zero, but their standard deviations were less than one.

Table 5.8. Descriptive Statistics and Reliability Estimates for the Performance Composites

Performance Composite	*n*	*Min*	*Max*	*M*	*SD*	r_{yy}
General Technical Proficiency (GTP)	768	-1.97	1.38	-0.01	0.52	.69
Achievement and Effort (w/ CSJT)	566	-1.64	2.23	0.02	0.52	.80
Achievement and Effort (w/o CSJT)	768	-1.73	2.37	-0.01	0.55	.77
Physical Fitness (PF)	768	-3.88	1.49	0.00	0.76	.92
Teamwork (TEAM)	768	-2.84	1.26	0.04	0.59	.35
Future Expected Performance (FXP)	768	-2.56	1.55	-0.01	0.65	.54

Reliability of Performance Composites

Reliabilities for the performance composites were estimated based on a variation on Mosier's (1943) formula for the reliability of a weighted composite. In this formula, true score variance is estimated by subtracting weighted residual error variances specific to each component of the composite from observed composite score variance. Most of the performance composites, however, included several components based on the ratings measures (e.g., AW COPRS or AW FX). Given the linked nature of the measurement design underlying the rating measures discussed earlier, the residual variances of these rating-based components were correlated. If not accounted for, these correlated errors would inflate the estimate of true composite variance based on Mosier's formula. Therefore, we modified the formula to account for correlated errors among performance rating components comprising the composites. This modified approach necessitated estimating the covariance matrix among the true scores underlying the components for each composite. To do this analysis, we fitted disaggregated CFA

models separately for each composite to estimate the corrected covariance matrix.[17] To address the problem of arbitrarily assigning raters mentioned earlier, we followed the same approach described in the previous sections. That is, we conducted analyses on 500 random samples and then averaged the estimates. The results were used in modified formulas to estimate reliabilities for the performance composites. Appendix B presents the formulas we used and their derivations. As shown in Table 5.8, the reliabilities of the performance composites varied greatly, ranging from .35 for the Teamwork composite to .92 for the Physical Fitness composite.

Composite Intercorrelations

Table 5.9 presents correlations among the performance composites. Values below the diagonal are observed correlations; those above the diagonal are factor correlations obtained from the performance models discussed in the two previous sections. Specifically, correlations among current performance factors were provided by the final cross-instrument current performance model (averaged across results from 500 modeling samples). Correlations between the Future Expected Performance composite and current performance composites were obtained from the current and future performance models described in the previous section. These factor-level correlations reflect the estimated correlations between the composites after accounting for the *attenuating* effects of unreliability, and the *inflationary* effects of having non-zero error covariances among the composites.[18]

Table 5.9. Intercorrelations of Composite Performance Criteria

Composite		1	2	3	4	5	6
1	General Technical Proficiency (GTP)	.	**.71**	.	**.26**	**.33**	**.72**
2	Achievement and Effort (w/ CSJT)	**.63**	.		**.36**	**.46**	**.62**
3	Achievement and Effort (w/o CSJT)	**.63**	**.95**	.	.	.	.
4	Physical Fitness (PF)	**.24**	**.25**	**.27**	.	-.10	**.31**
5	Teamwork (TEAM)	**.47**	**.48**	**.52**	**.08**	.	**.27**
6	Future Expected Performance (FXP)	**.73**	**.67**	**.69**	**.27**	**.49**	.

Note. n = 566-768 (for correlations below the diagonal). n = 370 (for correlations above the diagonal). Correlations below the diagonal reflect raw (unadjusted) correlations between observed composite scores. Correlations above the diagonal reflect mean corrected correlations (across 500 samples created for the modeling effort) between factors from the cross-instrument performance model. Statistically significant correlations are bolded ($p < .05$, two-tailed).

[17] Another option would have been to derive the corrected covariance matrix from the final cross-instrument CFA model. We did not adopt this option because it resulted in reliability estimates for non-rating components of the composites that were unrealistically low. For example, had we based the reliability estimate for the AWJKT on the CFA model, it would have been .02; recall that in Chapter 4 we reported the internal consistency reliability of the AWJKT to be .71. The reliabilities of the non-rating indicators in the CFA models (e.g., CSJT, AWJKT) are a direct function of their loading on their latent performance factor. Because loadings for these non-rating indicators were generally low (one exception was the loading of APFT scores on Physical Fitness), this produced extremely low reliability estimates based on this model. However, it is important to note that basing reliability estimates for the non-rating indicators on CFA models such as this may be problematic in that low "reliability" may be less an issue of high levels of measurement error (in the Classical Test Theory sense), and more of an issue of little saturation with variance from the latent performance factor of interest (an issue of construct-validity, or saturation of the performance factor with rating-specific variance). Therefore, we (a) fitted CFA models for each composite separately to generate corrected covariance matrixes for the components underlying each composite, (b) constrained the loadings for the non-rating indicators using the square root of the reliability reported for them in Chapter 4, and (c) constrained their residuals to be equal to 1.00 minus the reliability. The ratings parameters portion of the model was left to be freely estimated.

[18] Recall that the non-zero error covariances arise from having common raters across dimensions.

There was a large amount of variation across correlations among current performance composites. Corrected correlations ranged from -.10 to .71. Although the correlation between General Technical Proficiency and Achievement and Effort (w/CSJT) was sizable, it was not large enough to suggest that these composites are tapping the same construct, as they only shared 50% of their variance. With regard to relations between current performance composites and the future performance composite, corrected correlations ranged from .27 to .72. Although the correlations between Future Expected Performance and General Technical Proficiency and between Future Expected Performance and Achievement and Effort (w/o CSJT) were sizable, they were not so large as to suggest that future performance simply reflects Soldiers' current performance. Specifically, General Technical Proficiency and Future Expected Performance shared only 52% of their variance, whereas Achievement and Effort (w/o CSJT) and Future Expected Performance shared only 38% of their variance. Furthermore, on average, Future Expected Performance shared only 27% of its variance with the current performance composites. Thus, future performance appeared to be assessing a distinct construct that was not just current performance.

Subgroup Differences

Tables 5.10 and 5.11 show subgroup means on the performance composites by gender and race/ethnicity. In terms of gender, there were four statistically significant mean differences, and the effect sizes associated with those differences were small to moderate (0.33 to 0.49). Specifically, females had higher mean scores than males on both Achievement and Effort composites, as well as the Teamwork and Future Expected Performance composites. Small to moderate statistically significant mean differences were also found by race/ethnicity. For example, Black Soldiers scored lower than White Soldiers on both Achievement and Effort composites, General Technical Proficiency, and Future Expected Performance, whereas Hispanic Soldiers scored higher than did White non-Hispanic Soldiers on Achievement and Effort (w/o CSJT) and Teamwork.

Table 5.10. Performance Composite Scores by Gender

		Male		Female	
Composite	d_{FM}	*M*	*SD*	*M*	*SD*
General Technical Proficiency (GTP)	0.02	-0.01	0.52	0.00	0.52
Achievement and Effort (w/ CSJT)	**0.49**	-0.01	0.51	0.25	0.51
Achievement and Effort (w/o CSJT)	**0.41**	-0.03	0.55	0.20	0.55
Physical Fitness (PF)	-0.20	0.01	0.75	-0.14	0.82
Teamwork (TEAM)	**0.34**	0.02	0.59	0.22	0.61
Future Expected Performance (FXP)	**0.33**	-0.03	0.65	0.18	0.63

Note. n_{Male} = 500-692. n_{Female} = 65-75. d_{FW} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 5.11. Performance Composite Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
Composite	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
General Technical Proficiency (GTP)	**-0.45**	-0.08	0.04	0.53	-0.20	0.46	0.05	0.55	0.00	0.47
Achievement and Effort (w/ CSJT)	**-0.27**	0.12	0.05	0.51	-0.09	0.56	0.04	0.52	0.10	0.48
Achievement and Effort (w/o CSJT)	**-0.22**	**0.23**	0.01	0.54	-0.11	0.59	-0.01	0.56	0.11	0.50
Physical Fitness (PF)	0.01	0.10	-0.01	0.77	-0.01	0.77	-0.02	0.77	0.05	0.76
Teamwork (TEAM)	0.00	**0.24**	0.03	0.59	0.03	0.62	0.00	0.60	0.15	0.54
Future Expected Performance (FXP)	**-0.21**	0.10	0.02	0.67	-0.12	0.56	0.00	0.69	0.07	0.59

Note. n_{White} = 408-550. n_{Black} = 108-147. $n_{White\ Non\text{-}Hispanic}$ = 320 - 428. $n_{Hispanic}$ = 108-150. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-Non-Hispanic White mean difference. Effect sizes calculated as (mean of non-referent group – mean of referent group)/*SD* of referent group. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Relations between Performance Composites and Attitudinal Criteria

Table 5.12 shows correlations between the performance composites and the final attitudinal criteria identified in Chapter 3. In general, the attitudinal criteria appeared to be most related to the Achievement and Effort performance composite. Conceptually, this makes sense, as a common link between Achievement and Effort (which primarily reflects will-do performance) and attitudes may be work motivation. To the extent that performance composites such as General Technical Proficiency and Physical Fitness reflect can-do performance (arguably, primarily a function of knowledge, skills, and aptitudes), then the correlation with attitudes may be expected to be weaker. Another interesting pattern is that the performance criteria appeared to be far more related to Satisfaction, Perceived Fit with the Army, and Attrition Cognitions compared to Career Intentions and Future Army Affect. A key difference between the former attitudinal criteria and latter attitudinal criteria is that the latter tend to be future-oriented, and as such may be more a function of non-performance related factors (e.g., long-term goals, personal financial situation, reasons for joining the Army).

Table 5.12. Correlations between Performance Composites and Attitudinal Criteria

	Attitudinal Criterion				
Performance Composite	Satisfaction with the Army	Perceived Army Fit	Attrition Cognitions	Career Intentions	Future Army Affect
General Technical Proficiency (GTP)	**.11 (.08)**	**.24 (.18)**	**-.29 (-.19)**	-.01 (.00)	.02 (.01)
Achievement and Effort (w/ CSJT)	**.29 (.24)**	**.41 (.33)**	**-.36 (-.26)**	**.17 (.15)**	**.12 (.10)**
Achievement and Effort (w/o CSJT)	**.21 (.17)**	**.28 (.22)**	**-.28 (-.20)**	.09 (.07)	.00 (.00)
Physical Fitness (PF)	**.15 (.13)**	**.20 (.17)**	**-.22 (-.17)**	.07 (.07)	.03 (.03)
Teamwork (TEAM)	**.17 (.09)**	**.19 (.10)**	-.11 (-.06)	.03 (.02)	-.01 (.00)
Future Expected Performance (FXP)	**.13 (.09)**	**.23 (.15)**	**-.29 (-.18)**	.08 (.06)	.02 (.02)

Note. *n* = 534-707. Within each cell, correlations corrected for measurement error (in both measures) are shown first; raw correlations appear next in parentheses. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Summary

This chapter described results of modeling the Select21 performance domain, forming performance composites for use in subsequent chapters, and estimating the relationship between Select21 performance and attitudinal criteria. In general, the results of this modeling effort were quite similar to results of previous Army research. Specifically, latent performance factors that underlie the Select21 performance domain appear quite similar to those found in Project A (Campbell & Knapp, 2001). For example, like Project A, the Select21 performance model includes factors for General Technical Proficiency (similar in concept General Soldiering Proficiency factor in the five-factor model of first tour performance in Project A), Achievement and Effort (similar to the Effort and Leadership factor in Project A), and a Physical Fitness factor.

Although several factors are similar in *name* to those found in Project A, it is important to note that the models differ in some notable ways. For example, unlike Project A, we were unable to find evidence for an MOS-specific Core Technical Proficiency factor. The lack of evidence for such a factor in Select21 may simply reflect the fact that MOS-specific "hands-on" performance tests (e.g., work samples), and MOS-specific job knowledge tests were not included in Select21 as they were in Project A.[19] Another difference between the Select21 results and the first tour Project A results is that no evidence emerged in support of differentiating a Personal Discipline factor from Achievement and Effort. For example, whereas PFF Disciplinary Actions was associated with a Personal Discipline factor in Project A, here it appeared to provide just a negative indicator of Achievement and Effort. Lastly, a final key difference between models regards the General Technical Proficiency factor found in Select21 and General Soldiering Proficiency factor found in Project A.

In Project A, the General Soldering Proficiency factor consisted of a general hands-on performance test and job knowledge test, whereas in Select21, the General Technical Proficiency factor consisted of a job knowledge test, Army-wide performance rating scales, and a weapons qualification score. In Project A, the performance rating scales loaded primarily on Effort and Leadership, whereas in Select21, these rating scales loaded on several different factors (including General Technical Proficiency and Achievement and Effort). We hypothesize that the loading of performance ratings scales on both technical proficiency and effort-related factors in Select21 can be explained by differences in model fitting procedures used in Project A and Select21, as well as differences in the types of criteria examined. Models of the criterion space in Project A were fitted on aggregated ratings data. The correlated error arising from having common raters across the dimension would thus make it difficult to distinguish between rating scales that were designed to assess different performance constructs (e.g., General Soldiering Proficiency and Effort and Leadership). In Select21, we fitted all performance models on disaggregated ratings data to account for such error covariance, and this fact may have allowed us to make finer distinctions (relative to Project A) between ratings scales designed to assess different performance constructs. Differences in the types of criteria included in Project A and Select21 might also explain differences in the loadings of the rating scales. Specifically, whereas hands-on performance and job knowledge examined in Project A were primarily a function of declarative

[19] As discussed in Chapter 1, MOS-specific job knowledge tests were available for some, but not most, Soldiers in the Select21 sample.

knowledge (DK) and procedural knowledge and skill (PKS), performance ratings were a function of DK, PKS, and motivation (McCloy, Campbell, & Cudeck, 1994). Thus, had a hands-on performance test been available for Select21, its presence may have led to its clustering with the Army-wide job knowledge test under a general proficiency factor and to the clustering of the performance ratings scales under an effort-related performance factor (reflecting the scales' links to motivation).

Based on the results of the modeling effort we formed performance composites, all of which appear to have adequate discriminant validity, and most of which appear to have adequate reliability. The estimated reliability of the Teamwork (.35) and Future Expected Performance (.54) composites were quite low, particularly given they reflect the average across multiple raters (i.e., they are not single-rater reliability estimates). The low reliability of the composites can be traced back to the low interrater reliability found for individual performance dimensions that underlie these composites (presented in Chapter 4).

Examination of the pattern of relations among performance and attitudinal criteria revealed some findings of note. For example, the Achievement and Effort performance composite was the performance composite most strongly related to current-focused attitudes such as Satisfaction with the Army and Perceived Army Fit. Additionally the performance criteria in general appeared to hold stronger relations with the current-focused attitudinal criteria, compared to the more distal future-oriented attitudes regarding Career Intentions and Future Army Affect.

PART 3: INDIVIDUAL PREDICTORS AND BIVARIATE VALIDITY RESULTS

CHAPTER 6: PREDICTOR MEASURE VALIDATION METHODS AND ARMED SERVICES VOCATIONAL APTITUDE BATTERY RESULTS

Teresa Russell, Huy Le, and Dan Putka
HumRRO

Overview

This chapter has two purposes. First, because it is the first of a series of chapters reporting results for predictors, it explains the methods used in all of the predictor chapters to estimate validity, incremental validity, subgroup differences, and differential prediction. Second, it reports the results of psychometric analyses of selected scores from the Armed Services Vocational Aptitude Battery (ASVAB) using the full Select21 concurrent validation sample—scores that will be used in analyses reported in later chapters. For each methodology (e.g., validity estimation), we describe the method and then provide the ASVAB results as an illustration before turning to the next methodology.

ASVAB Background

The ASVAB is a differential aptitude battery, philosophically a descendent of Thurstone's (1938) research to define primary mental abilities. The content of the ASVAB stems from modifications of the Army General Classification Test (AGCT) and the Navy General Classification Test (NGCT) that were used during World War II (Schratz & Ree, 1989). Separate batteries were used until the late-1960s when the Services developed a joint testing program. The resulting multiple-aptitude, group-administered ASVAB is now the primary enlisted personnel selection test used by the military.

Numerous validity studies have shown that the ASVAB is a valid predictor of training performance (e.g., Ree & Earles, 1991; Welsh, Kucinkas, & Curran, 1990), job performance in the first tour (e.g., McHenry, Hough, Toquam, Hanson, & Ashworth, 1990), and job performance in the second tour (Campbell & Johnson, 1992; Oppler, Peterson, & Rose, 1996).

The current version of the ASVAB contains the following nine subtests:

- General Science (GS)
- Arithmetic Reasoning (AR)
- Word Knowledge (WK)
- Paragraph Comprehension (PC)
- Auto and Shop Information (AS)
- Math Knowledge (MK)
- Mechanical Comprehension (MC)
- Electronics Information (EI)
- Assembling Objects (AO)

All of the subtests except AO are used in the Army's selection and classification composites. AO is an experimental spatial ability test. For that reason, we are especially interested in conducting analyses with AO in this chapter (referred to as "Spatial" in the rest of this report). Other than Spatial, we will not focus on any of the individual subtests in this research.

Two ASVAB composite scores merit special attention in Select21—the Armed Forces Qualification Test (AFQT) and ASVAB Technical. AFQT is clearly important because it is the composite used by the Army for selection purposes. Technical is important because it contains the technical information tests that supplement the broader verbal and math tests on the ASVAB. The formulas for AFQT and Technical are as follows:

AFQT = AR + MK + 2VE, where VE = WK + PC.
Technical = AS + MC + EI.

The analyses use AFQT and Spatial scores obtained from operational personnel data files. AFQT is a percentile score. The Spatial (i.e., Assembling Objects) score and the other subtest scores are standardized scores (M = 50, SD = 10). We computed the Technical composite by simply adding the subtest scores from the operational data files together.

Table 6.1 provides the means, standard deviations, and correlations between the three scores of interest (AFQT, Technical, and Spatial) and the eight operational ASVAB subtests. The concurrent validation (CV) sample *SD*s illustrate the effect of range restriction in the sample because they are lower than the population *SD*s which are approximately 10. The CV sample correlations (uncorrected) appear below the diagonal and norming study subtest correlations appear above the diagonal. Subtest reliability estimates appear on the diagonal. As shown, the sample specific correlations were notably lower than their unrestricted population counterparts. All of the subtest correlations in the CV sample were significant ($p < .01$, one-tailed) except the remarkably low correlation ($r = .01$) between MK and AS, likely due in part to the range restriction on AFQT (i.e., MK is included in AFQT). But, as shown, the MK/AS correlation was also low in the norming study population ($r = .24$).

Table 6.1 Descriptive Statistics and Correlations for ASVAB Scores in the Full CV Sample

	Select21 Full CV Sample		Correlations									
Score	*M*	*SD*	GS	AR	WK	PC	AS	MK	MC	EI	AFQT	T
General Science (GS)	51.99	7.49	*.84*	.72	.80	.72	.52	.69	.68	.70		
Arithmetic Reasoning (AR)	51.63	7.21	**.47**	*.87*	.67	.72	.42	.80	.65	.60		
Word Knowledge (WK)	52.51	5.75	**.67**	**.37**	*.89*	.76	.43	.61	.58	.61		
Paragraph Comprehension (PC)	52.76	6.41	**.52**	**.43**	**.56**	*.75*	.35	.68	.59	.55		
Auto and Shop Information (AS)	48.76	8.02	**.44**	**.29**	**.38**	**.27**	*.83*	.24	.67	.72		
Math Knowledge (MK)	53.97	6.82	**.32**	**.60**	**.19**	**.30**	.01	*.84*	.55	.48		
Mechanical Comprehension (MC)	52.16	8.36	**.53**	**.51**	**.42**	**.39**	**.57**	**.28**	*.79*	.71		
Electronics Information (EI)	50.60	7.89	**.53**	**.35**	**.43**	**.32**	**.57**	**.19**	**.52**	*.72*		
AFQT	57.33	18.15	**.65**	**.82**	**.70**	**.67**	**.30**	**.72**	**.53**	**.46**		
Technical (T)	151.53	2.33	**.60**	**.46**	**.49**	**.39**	**.85**	**.19**	**.84**	**.83**	**.51**	
Spatial (S)	52.53	8.73	**.39**	**.40**	**.23**	**.28**	**.31**	**.35**	**.55**	**.29**	**.38**	**.46**

Note. Select21 full concurrent validation (CV) sample n = 771 for all subtests and correlations except those involving Spatial. n Spatial = 577. Select21 CV sample correlations appear below the diagonal. Correlations that are significant at the $p < .01$ (one-tailed) level are in bold. Correlations between ASVAB subtests in the Profile of American Youth 1997 (PAY97) population appear above the diagonal. Alternate forms reliabilities (Forms 10a and 11a) appear in italics on the diagonal (Palmer, Hartke, Ree, Welsh, & Valentine, 1988).

Zero-Order Criterion-Related Validity Estimates

Method

All the chapters in this report use the following three-step method of computing zero-order criterion-related validity estimates for a predictor score:

1. Compute zero-order validity estimates by correlating each criterion score with the predictor score.
2. Correct the zero-order validity estimates for criterion unreliability (Hunter & Schmidt, 1990).
3. Correct the zero-order validity estimates from Step 2 for range restriction on AFQT (direct range restriction in case of AFQT and indirect range restriction in cases of Technical and Spatial scales, Lord & Novick, 1968). AFQT is a percentile score and, as such, its scores in the population have a uniform distribution (i.e., rectangular). The formula for the population variance for a rectangular distribution is:

$$var\_rect = (b - a)^2/12,$$

where b and a are the endpoints of the uniform distribution.

Replacing b and a with 100 and 1, respectively, yields var_rect = 816.75, or an *SD* of 28.58.

All validity estimates were computed for the full CV sample.

ASVAB Results

The general format for the zero-order validity results appears in Table 6.2. The raw, zero-order validity estimates appear in the upper half of the zero-order validity table. The zero-order validity coefficients corrected for criterion unreliability and for range restriction on AFQT appear in the lower half of the zero-order validity table. The five performance and five attitudinal criterion composites are described in Chapters 3-5. In short, they are:

Performance Criteria:

- GTP—General Technical Proficiency includes Army-Wide job knowledge test scores, the Personnel File Form Weapons Qualification score, and performance ratings on technical dimensions.
- AE—Achievement and Effort includes performance ratings, and in all the chapters of this report except one, it includes scores on the Criterion Situational Judgment Test (CSJT). Analyses for the Predictor SJT (PSJT) use a version of the AE composite without the CSJT.
- PF—Physical Fitness includes the Army Physical Fitness Test score and performance ratings.
- TEAM—Teamwork includes performance ratings.
- FXP—Future Predicted Performance includes future expected performance ratings.

Attitudinal Criteria:

- ASat—Satisfaction with the Army from the Army Life Survey (ALS).
- AFit—Perceived Army Fit from the ALS.
- CInt—Career Intentions from the ALS.
- ACog—Attrition Cognitions from the ALS.
- FAA—Future Army Affect from the Future Army Life Survey (FALS).

Table 6.2. Uncorrected and Corrected Zero-Order Validities for ASVAB Test Scores

Predictor Scale	Performance Criteria					Attitudinal Criteria				
	GTP	AE	PF	TEAM	FXP	ASat	AFit	CInt	ACog	FAA
	Uncorrected Validity Estimates									
AFQT	**.30**	**.16**	.00	**.06**	**.17**	-.01	.00	**-.07**	**-.12**	-.05
Spatial	**.21**	**.11**	.04	.01	**.15**	-.01	.03	-.02	**-.07**	.03
Technical	**.29**	**.09**	-.04	.05	**.11**	-.01	.00	-.05	**-.09**	.05
	Corrected Validity Estimates									
AFQT	.52	.28	.00	.16	.35	-.02	.01	-.11	-.23	-.08
Spatial	.38	.20	.04	.07	.29	-.02	.03	-.06	-.15	.01
Technical	.48	.20	-.04	.13	.27	-.02	.00	-.09	-.18	.01

Note. n = 414 - 739. Statistically significant correlations are bolded ($p < .05$, two-tailed). Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

As shown in Table 6.2, AFQT, Spatial, and Technical yielded significant correlations with General Technical Proficiency, Achievement and Effort, and Future Expected Performance scores. They were not strong predictors of Physical Fitness and Teamwork performance. With regard to attitudinal variables, higher ASVAB scores appeared to be related to having fewer thoughts about attriting from the Army and lower intentions to reenlist. While seemingly counterintuitive, this finding is consistent with prior research (e.g., Strickland, 2005). Apparently, Soldiers with higher AFQT scores are less likely to plan to make the Army a career, but are more likely to plan to honor their initial enlistment commitment, than Soldiers with lower AFQT scores.

These results appear to be in line with other ASVAB research. Unfortunately, most reported ASVAB validities are based on correlations with training grades instead of job performance. Corrected zero-order correlations between AFQT and final school grades from training are typically in the upper .60s or lower .70s (c.f. Oppler, Russell, Rosse, Keil, Meiman, & Welsh, 1997; Ree & Earles, 1991; Welsh, Kucinkas, et al., 1990). When job performance criteria have been used, the ASVAB scores have not been formulated like those in the current and past ARI research. For example, In Project A, corrected/adjusted validity estimates for the full ASVAB were .71 for predicting Core Technical Proficiency, .75 for predicting General Soldiering Proficiency, and .40 for Effort and Leadership (Campbell & Knapp, 2001).

Incremental Validity Estimates

Method

Incremental validity is an estimate of the change in the multiple correlation (ΔR) when a new predictor is added to a regression equation. New predictors that add validity beyond that already afforded by AFQT are more likely to prove useful for selection purposes. Therefore, we computed raw and corrected/adjusted incremental validities for each predictor in this report.

The following steps were used to compute the raw incremental validity estimates for each predictor-criterion combination:

- Compute the correlation (*R)* for AFQT alone by regressing each criterion on AFQT.
- Compute the multiple *R* for AFQT and the new predictor by regressing each criterion on AFQT and the new predictor (i.e., AFQT + New Predictor).
- Compute the uncorrected incremental validity estimates (over AFQT) by subtracting the uncorrected correlation for model with AFQT only obtained from Step 1 from the uncorrected multiple *R*(AFQT + New Predictor) obtained from Step 2.

Calculating corrected incremental validity estimates involved a few more steps. Those steps included the following:

- Compute the correlations among the new predictor, AFQT, and the criterion.
- Correct the correlations between (a) AFQT and the criterion and (b) the new predictor and the criterion for criterion unreliability.
- Correct the resulting *R*s for range restriction:
    - Correct the resulting correlations between AFQT and the predictor and the criterion for direct range restriction on AFQT (i.e., range restriction due to explicit selection on the AFQT; Lord & Novick, 1968) to the unrestricted AFQT *SD* = 28.58.[20]
    - Correct the resulting correlation between the predictor and the criterion for indirect range restriction (i.e., indirect range restriction on the predictor due to explicit selection on AFQT).
    - Correct the multiple *R*(AFQT + Predictor) for indirect range restriction.
        - Generate a corrected 3 x 3 correlation matrix consisting of corrected bivariate correlations between the criterion, AFQT, and the predictor obtained in the previous steps (using only those Soldiers who have all three scores). Regress the criterion on AFQT and the predictor using this corrected matrix as input to arrive at a corrected estimate for multiple *R*.
    - Adjust the corrected *R*(AFQT + Predictor) for shrinkage using Rozeboom's (1978) Formula 8.
- Compute the corrected and adjusted incremental validity estimates (over AFQT) by subtracting the corrected *R*(AFQT) from the corrected and adjusted multiple *R*(AFQT + Predictor).

[20] The AFQT scores analyzed here are expressed as a percentile scores normed on the youth population. By definition, percentile scores have a mean of 50 and an *SD* of 28.58 in the norming population.

ASVAB Results

The general format for the incremental validity results appears in Table 6.3. The uncorrected incremental validity estimates appear in the upper half of the table with significant incremental validity estimates in bold ($p < .05$, two-tailed). The corrected and adjusted incremental validity estimates appear in the lower half.

One of the more notable results was that the corrected/adjusted incremental validity coefficients at the bottom of the page are generally lower than the uncorrected ones. There are two reasons for this finding. First, there is direct range restriction on AFQT. When *R*(AFQT) is corrected for range restriction, it increases (e.g., from .30 to .52 for predicting General Technical Proficiency) making it much more difficult to show Δ*R*. Second, the adjustment for shrinkage also lowers the corrected incremental validities. This reduction of corrected incremental validities was observed for most of the predictors in this report.

Table 6.3. Incremental Validity Estimates for ASVAB Test Scores

	Performance Criteria					Attitudinal Criteria				
Predictor Scale	GTP	AE	PF	TEAM	FXP	ASat	AFit	CInt	ACog	FAA
	Uncorrected Validity Estimates									
AFQT	**.30**	**.16**	.00	**.06**	**.17**	-.01	.00	**-.07**	**-.12**	-.05
AFQT + Spatial	**.02**	.01	.05	.00	**.02**	.00	.03	.00	.00	.03
AFQT + Technical	**.04**	.00	.05	.00	.00	.00	.00	.00	.00	**.05**
AFQT + Technical + Spatial	**.07**	.01	.02	.01	**.04**	.02	.04	.00	.02	**.09**
	Corrected Validity Estimates									
AFQT	.52	.28	.00	.16	.35	-.02	.01	-.11	-.23	-.08
AFQT + Spatial	.01	.00	.00	.00	.01	.00	.00	.00	.00	.00
AFQT + Technical	.02	.00	.00	.00	.00	.00	.00	.00	.00	.01
AFQT + Technical + Spatial	.05	.00	.00	.00	.02	.00	.00	.00	.00	.05

Note. $n = 414 - 739$. Cell values for the AFQT represent zero-order correlations between AFQT and the given criterion (shown for reference). Uncorrected incremental estimates reflect the difference between the multiple *R* obtained when regressing the criterion on both the given composite and AFQT versus the *R* obtained when regressing the criterion only on the AFQT. Statistically significant incremental validity coefficients are bolded ($p < .05$, one-tailed). Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

As shown, the Spatial and Technical composites (alone and together) provided incremental validity over AFQT for the prediction of General Technical Proficiency even after correction and adjustment. It is important to note that the Spatial score provided incremental validity beyond that provided by the AFQT along with the Technical score (i.e., [corrected $R_{\text{AFQT + Technical + Spatial}} = .05$] minus [corrected $R_{\text{AFQT+ Technical}} = .02$] = .03 Δ*R*). This finding suggests that Spatial could be a useful predictor beyond the ASVAB, not just beyond AFQT.

Although the Spatial and Technical scores would not typically be expected to predict attitudinal criteria, there appeared to be some incremental validity for predicting attitudes about the future Army. Note that AFQT was negatively correlated with Future Army Affect (Table 6.2), while the other Spatial and Technical scores were slightly positively correlated with it.

Subgroup Differences

Method

This chapter and subsequent chapters report subgroup difference effect sizes to indicate the magnitude of the difference between subgroups' scores. Effect sizes are standardized mean difference scores and are thus interpreted in standard deviation units. The subgroup difference effect size formula used in this report is as follows:

d = (mean of non-referent group — mean of referent group)/*SD* of the referent group.

The referent group is the group that does not have special protections under relevant employment laws (i.e., males and Whites). Referent groups are listed second in the effect size subscript.

ASVAB Results

As shown in Table 6.4, there was typically little or no difference between males' and females' scores on AFQT. The difference in the Select21 sample was relatively small, and the difference in the ASVAB norming population was even smaller (PAY80; U.S. Department of Defense, 1982).[21] In contrast, there were relatively large differences between male and female subgroup scores on the Technical composite for both the Select21 and norming samples. Females scored approximately one-third of an *SD* lower than males on Spatial in the Select21 sample. The Spatial test was not administered in PAY80.

Table 6.4. ASVAB Scores by Gender

Score	PAY80†	S21	Male		Female	
	d_{FM}	d_{FM}	*M*	*SD*	*M*	*SD*
AFQT	-.05	-.15	57.62	18.26	54.88	17.10
Spatial	--	**-.30**	52.82	8.58	50.20	9.65
Technical	-.95	**-1.15**	153.91	19.42	131.51	16.50

Note. S21 n_{Male} = 513-689, S21 n_{Female} = 64-82, d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of non-referent group – the mean of referent group)/*SD* of the referent group. Referent groups (e.g., Males) are listed second in the effect size subscript. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed). A positive effect size indicates that on average the non-referent group performs better in the tests.
†Profile of American Youth (PAY80) results adapted from U.S. Department of Defense (1982). *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery (ASVAB).* Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics). PAY80 *d* for the technical score is the mean of the effect sizes for MC, EI, and AS. The Spatial test was not administered in PAY80.

As shown in Table 6.5, race/ethnic subgroup differences in AFQT scores were substantially smaller in the Select21 sample than they were in the 1980 norming population, suggesting fairly large differences in these samples. Of course, the Select21 sample was range restricted on AFQT since Soldiers were selected on this measure; therefore, much of the difference is likely due to range restriction. Effect sizes for the Select21 sample were also smaller for the Technical composite but not to the extent of the AFQT.

[21] Subtest scores for the PAY97 norming population have not yet been published.

Table 6.5. ASVAB Scores by Race/Ethnic Group

	PAY80†		S21		White		Black		White Non-Hispanic		Hispanic	
Score	d_{BW}	d_{HW}	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
AFQT	-1.21	-.94	**-.46**	**-.48**	59.70	18.27	51.30	15.85	61.53	18.10	52.79	17.79
Spatial	--	--	**-.46**	-.13	53.29	8.52	49.40	9.03	53.38	8.22	52.30	9.72
Technical	-1.22	-.86	**-.98**	**-.80**	156.21	19.28	137.38	16.10	159.01	17.45	145.03	21.14

Note. S21n_{White} = 415-549, S21 n_{Black} = 113-151, S21$n_{White\ Non\text{-}Hipanic}$ = 328-425, S21$n_{Hispanic}$ = 107-154. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of non-referent group – mean of referent group)/*SD* of referent group. Referent groups (e.g., White) are listed second in the effect size subscript. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed). †Profile of American Youth (PAY80) results adapted from U.S. Department of Defense (1982). *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery (ASVAB).* Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics). *d* for the technical score is the mean of the effect sizes for MC, EI, and AS. The Spatial test was not available for PAY80.

Differential Prediction

Method

An important aspect of any validation effort is to investigate potential bias in a measure. The professionally accepted method of assessing bias is Cleary's (1968) differential prediction model (AERA, APA, NCME, 1999; SIOP, 2003). According to that model, a measure is not biased if regression lines (using scores on the measure to predict performance) for the subgroups are not significantly different with regard to the standard errors of estimate (SEE), slopes, and intercepts. The SEE, slope, and intercept are illustrated in Figure 6.1. The SEE is an index of the amount of error in prediction—the scatter of observed scores around the predicted score. SEE differences are usually not significant and rarely tested. The tendency is to be permissive with respect to violations of SEE equality (Humphreys, 1986).

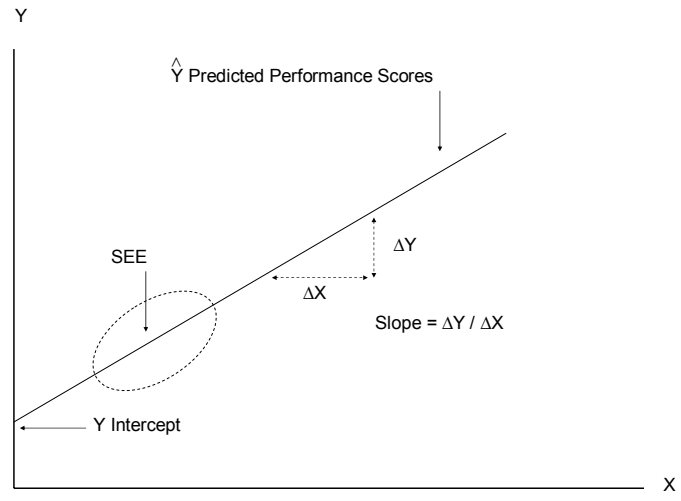


Figure 6.1. Differential prediction concepts.

Slope and intercept differences can be evaluated by fitting a moderated multiple regression (MMR) model to the data. MMR involves sequential comparison of regression models, testing first for differences in slopes, then for differences in intercepts (Bartlett, Bobko, Mosier, & Hannan, 1978).

Caveats

In reviewing differential prediction results throughout this report, there are at least three caveats to keep in mind. First, our sample sizes for some of the non-referent groups were smaller that what is desirable for MMR analyses. When sample sizes are small, MMR results are not stable and the slope test, in particular, lacks power (Linn, 1994). This is particularly of concern for the gender-related Select21 MMR analyses since the number of females in the sample was relatively small. Second, differential prediction results should be interpreted within the context of the overall validity between the predictor and the criterion for the entire sample. That is, if the predictor score is not a valid predictor of the criterion, slope and intercept differences for that predictor-criterion combination may not be of practical concern. For example, ASVAB test scores were not very useful predictors of attitudinal criteria (Tables 6.2 and 6.3), nor was the ASVAB developed for this purpose. It was developed to predict training and job performance. Findings of differential prediction of ASVAB scores for attitudinal variables may not be of much concern. Third, whenever regression models are used, it is important to remember that other variable(s) excluded from the analyses could impact the relationship between predictor and criterion.

ASVAB Results

Slope Differences for Gender and Race/Ethnicity

Slope bias reflects differences in the slopes associated with the measure in regression lines fit for each subgroup as shown in Figure 6.2. Slope bias suggests that the measure is more predictive of performance for one subgroup than another. The slope test lacks power to detect slope differences for the typical sample sizes in studies (Linn, 1994).

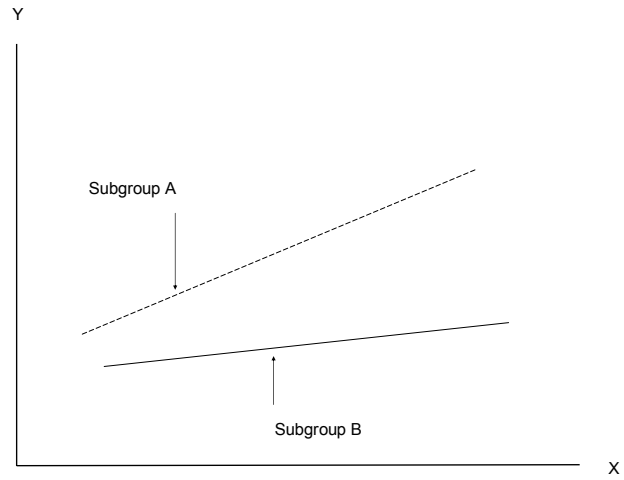


Figure 6.2. Subgroup slope differences.

In the context of MMR analysis, slope bias is evidenced by a significant interaction between the score on the measure and subgroup membership. This report uses one general format for reporting differential prediction results, as shown in Table 6.6. Slope differences are reported under the "AFQT *b,*" "Spatial *b,*" and "Technical *b,*" columns. For the referent group (i.e., males, Whites), these values are simply the unstandardized regression weights associated with the measure's score. For the non-referent group (e.g., females) these values are the sum of the unstandardized regression weights associated with the score, and the cross-product term (score x subgroup) from the MMR analyses. Regression weights are bolded if the score-by-gender interaction term (i.e., slope difference) was statistically significant.

For example, Table 6.6 shows differential prediction results by gender for the ASVAB scores. One slope difference out of 30 slope tests conducted was significant. It was for regressing the Achievement and Effort criterion score on AFQT. Since females had a significantly steeper slope than males, the regression weights were bolded under the "AFQT *b*" column. The values under the "*r* by Gender" columns in Table 6.6 contain uncorrected zero-order correlations between ASVAB scores and criteria for each gender separately. As shown, AFQT was a valid predictor of Achievement and Effort for both groups, though the validity for females was higher.

Table 6.7 reports results of the differential prediction analyses comparing White and Black Soldiers. As shown by the bolded values, three of 30 slope tests were significant. Results for ethnic subgroups (White, Hispanic) are shown in Table 6.8, which also show three of 30 slope tests yielding significant differences.

Intercept Differences for Gender and Race/Ethnicity

Intercept bias reflects differences in the intercepts of regression lines fitted for each subgroup as shown in Figure 6.3. Intercept bias suggests that the measure would underpredict performance for one group relative to another if a common regression line were used to predict performance. If a slope difference is significant, intercept differences are more complicated; the subgroup's performance might be underpredicted in some parts of the distribution and overpredicted in others. In the cognitive domain, when intercept differences are significant, they usually indicate overprediction of the protected group (Bartlett et al., 1978; Hunter, Schmidt, & Rauschenberger, 1977; Schmidt, Pearlman, & Hunter, 1980). In other domains, there has not been sufficient research to support general conclusions (SIOP, 2003).

Table 6.6. Differential Prediction Results for ASVAB Scores by Gender

	AFQT					Spatial					Technical				
	Gender	AFQT *b*		*r* by Gender		Gender	Spatial *b*		*r* by Gender		Gender	Technical *b*		*r* by Gender	
Criterion	*b*	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F
General Technical Proficiency	.03	.15	.18	**.30**	**.33**	.03	.11	.09	**.21**	.20	.17	.17	.15	**.31**	.23
Achievement and Effort	**.26**	**.07**	**.22**	**.15**	**.38**	**.23**	.05	.10	**.11**	.23	**.37**	.08	.11	**.16**	.18
Physical Fitness	-.14	.00	.01	.00	.01	-.05	.01	.18	.01	.24	-.22	-.05	-.07	-.07	-.07
Teamwork	**.20**	.04	.03	.07	.05	**.25**	.01	.05	.01	.12	**.22**	.05	.03	**.09**	.03
Future Expected Performance	**.23**	.11	.18	**.17**	**.27**	**.25**	.09	.20	**.14**	**.34**	**.36**	.10	.14	**.15**	.18
Satisfaction with the Army	-.18	-.02	.02	-.02	.02	**-.22**	-.04	.10	-.05	.15	-.08	-.04	.11	-.05	.12
Perceived Army Fit	-.04	.01	-.03	.01	-.04	-.01	.01	.14	.01	.19	.11	-.02	.14	-.02	.14
Attrition Cognitions	**.32**	-.13	.01	**-.13**	.01	**.32**	-.04	-.19	-.04	-.20	.32	-.06	.01	-.06	.01
Career Intentions	-.02	-.06	-.21	-.05	-.18	-.04	-.05	.09	-.04	.08	-.10	-.06	-.10	-.05	-.07
Future Army Affect	**-.29**	-.03	-.19	-.03	**-.21**	**-.33**	.04	-.08	.04	-.11	**-.45**	.04	-.19	.04	-.18

Note. $n_{Regression}$ = 414-739. n_{Male} = 363-665. n_{Female} = 51-79. Gender *b* = Unstandardized regression weight for gender (0 = male, 1 = female). ASVAB score *b* = Unstandardized regression weight for the given ASVAB score for males and females. *r* by Gender = Correlation between the given ASVAB score and the given criterion for each gender. Regression weights for males and females are bolded if the score-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 6.7. Differential Prediction Results for ASVAB Scores by Race

	AFQT					Spatial					Technical				
		AFQT *b*		*r* by Race			Spatial *b*		*r* by Race			Technical *b*		*r* by Race	
Criterion	Race *b*	W	B	W	B	Race *b*	W	B	W	B	Race *b*	W	B	W	B
General Technical Proficiency	**-.16**	.16	.09	**.30**	**.17**	**-.19**	.08	.10	**.16**	**.25**	-.11	.16	.09	**.29**	**.17**
Achievement and Effort	-.10	.09	.01	**.18**	.01	**-.18**	.04	.06	.09	.12	-.13	.04	-.04	.07	-.05
Physical Fitness	-.02	**.04**	**-.14**	.05	-.16	.02	.08	-.08	.10	-.11	-.01	-.01	-.04	-.02	-.04
Teamwork	.05	.04	.05	.07	.07	.06	.02	.01	.03	.02	.09	.02	.10	.03	.13
Future Expected Performance	-.10	**.14**	**-.01**	**.20**	-.01	**-.18**	.10	.02	**.15**	.04	-.11	.09	-.01	**.13**	-.02
Satisfaction with the Army	-.04	.00	-.03	.00	-.04	.02	-.04	.12	-.05	.16	.00	-.02	.05	-.03	.05
Perceived Army Fit	-.10	.00	-.02	.00	-.02	-.08	.00	.10	.00	.14	-.15	-.01	-.07	-.01	-.07
Career Intentions	.02	-.05	-.20	-.05	-.15	-.02	-.02	-.07	-.02	-.07	-.16	**-.02**	**-.36**	-.02	**-.24**
Attrition Cognitions	**.30**	-.10	-.05	**-.10**	-.04	**.27**	-.03	-.14	-.03	-.14	**.41**	-.05	.12	-.05	.09
Future Army Affect	-.15	-.05	.01	-.06	.01	-.15	.03	.02	.03	.02	-.07	.03	.10	.03	.08

Note. $n_{Regression}$ = 380-671. n_{White} = 302-530. n_{Black} = 78-141. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black). ASVAB score *b* = Unstandardized regression weight for the given ASVAB score for Whites and Blacks. *r* by Race = Correlation between the given ASVAB score and the given criterion for each race. Regression weights for Whites and Blacks are bolded if the score-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 6.8. Differential Prediction Results for ASVAB Scores by Ethnic Group

	AFQT					Spatial					Technical				
	Ethnicity	AFQT *b*		*r* by Ethnicity		Ethnicity	Spatial *b*		*r* by Ethnicity		Ethnicity	Technical *b*		*r* by Ethnicity	
Criterion	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H
General Technical Proficiency	.02	.17	.12	**.31**	**.25**	**-.13**	.11	.04	**.19**	.08	.06	.19	.13	**.30**	**.28**
Achievement and Effort	.11	.11	.05	**.22**	.10	.00	**.09**	**-.04**	**.18**	-.10	.11	.07	.05	**.12**	.10
Physical Fitness	.07	**.07**	**-.08**	.09	-.11	.06	.11	.00	**.14**	.01	.06	.02	-.10	.03	-.14
Teamwork	**.17**	.03	.07	.06	.13	.09	.02	.04	.03	.07	**.17**	.04	.05	.05	.10
Future Expected Performance	.13	.17	.04	**.24**	.07	-.01	.14	.05	**.20**	.08	**.14**	.15	.04	**.19**	.07
Satisfaction with the Army	.12	-.02	.02	-.03	.02	**.19**	-.04	-.01	-.05	-.01	.12	-.03	.05	-.03	.07
Perceived Army Fit	.12	-.03	.04	-.03	.05	.11	-.02	.06	-.03	.10	.12	-.01	.03	-.01	.04
Career Intentions	.00	-.09	-.06	-.08	-.06	-.06	**-.09**	**.16**	-.07	.18	.04	-.05	.05	-.04	.06
Attrition Cognitions	-.06	-.10	-.10	**-.11**	-.11	.01	-.04	-.03	-.04	-.04	-.05	-.11	-.01	-.10	-.01
Future Army Affect	**.21**	-.06	.01	-.07	.01	**.22**	.03	.05	.03	.06	**.27**	.04	.10	.04	.12

Note. $n_{Regression}$ = 315-558. $n_{White,non-Hispanic}$ = 236-409. $n_{Hispanic}$ = 79-149. Ethnicity *b* = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). ASVAB score *b* = Unstandardized regression weight for the given ASVAB score for White non-Hispanics and Hispanics. *r* by Ethnicity = Correlation between the given ASVAB score and the given criterion for each ethnic group. Regression weights for White non-Hispanics and Hispanics are bolded if the score-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

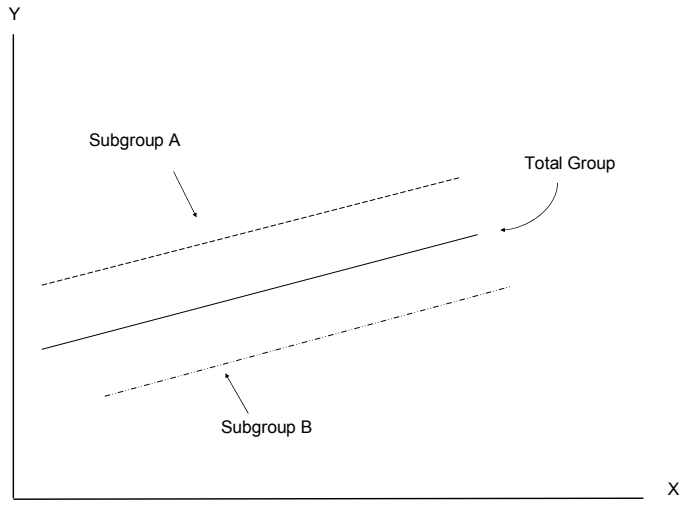


Figure 6.3. Subgroup intercept differences.

In the context of MMR analysis, an intercept difference is evidenced by a significant main effect for subgroup membership (e.g., gender, race). In Table 6.6, values reported under the "Gender *b*" columns are the unstandardized regression weights (*b*) associated with gender from the MMR analyses. These values reflect the predicted difference between females and males' raw criterion scores at the mean ASVAB score (across genders). Significant regression weights are bolded. A positive value indicates underprediction because the non-referent group (e.g., females) intercept is higher than the referent group intercept—the non-referent groups' scores would be underpredicted by the regression line for the entire sample.

For example, Table 6.6 shows that 15 of the 30 intercept tests conducted were significant for the gender comparisons. All three ASVAB scores underpredicted females' Achievement and Effort, Teamwork, and Future Expected Performance scores. MMR results for gender comparisons have not been widely reported in the industrial/organizational research literature making it difficult to draw sweeping generalizations. However, several studies have noted underprediction of women's grades in college based on college entrance exams (Gamache & Novick, 1985; Linn, 1973, 1982). Dunbar and Novick (1988) compared regressions for men and women in nine clerical Marine Corps jobs. ASVAB composites underpredicted females' final school grades in all nine jobs. But, underprediction of females' performance does not appear to be "the" common finding. Roberts and Skinner (1996) found that three cognitive test composites overpredicted women's grades in Officer training School. The same three composites yielded one slope difference, one overprediction, and one "no-difference" result against a ratings criterion. Meta-analytic or systematic reviews of the gender-related differential prediction literature are needed to better understand the findings.

As shown by the bolded values in Table 6.7, seven of the 30 intercept tests for race differences were significant. Intercept differences, when they appeared, suggested that the ASVAB overpredicted the performance of Black Soldiers on the performance criteria. For ethnic subgroups (Table 6.8), eight of 30 intercept tests were significant. In general, AFQT and the Technical composite tended to underpredict Hispanics' Teamwork performance. The Spatial score tended to overpredict General Technical Proficiency for Hispanics.

With regard to attitudinal criteria, ASVAB scores tended to (a) overpredict females' satisfaction with the Army and attitudes about the future Army, (b) underpredict Black Soldiers' attrition cognitions, and (c) underpredict Hispanic Soldiers' attitudes about the future Army.

Summary

This chapter (a) reported the results of psychometric analyses of selected scores from the ASVAB using the full Select21 concurrent validation sample and (b) explained the methods used in all of the remaining predictor chapters to estimate validity, incremental validity, subgroup differences, and differential prediction.

Review of ASVAB Results

AFQT, Spatial, and Technical scores yielded significant correlations with General Technical Proficiency, Achievement and Effort, and Future Expected Performance scores. They were not strong predictors of Physical Fitness and Teamwork performance. In contrast, the ASVAB score yielded a few significant, but relatively smaller, correlations with attitudinal variables. Higher AFQT scores tended to predict having fewer thoughts about leaving the Army prior to the end of the enlistment contract, but lower intentions to make the Army a career.

Some of the differences between mean ASVAB scores for subgroups were significant. The gender difference on AFQT was not significant. However, significant differences of about one-third *SD* on Spatial and over one *SD* on Technical did occur, with males receiving the higher scores on both. Race differences were significant for all three scores. The differences were about one-half *SD* on AFQT and on Spatial and one *SD* on Technical, with Whites receiving the higher scores. For the ethnic comparison, White Non-Hispanics received significantly higher scores by about one-half *SD* on AFQT and over three-quarters of an *SD* on Technical.

Differential prediction analyses indicated that gender comparisons tended to yield significant differences more frequently than race or ethnicity. That is, 15 out of 30 intercept tests and one out of 30 slope tests were significant for the gender comparisons. Seven of the 30 intercept tests and 3 of 30 slope tests were significant for the race comparison. Eight of 30 intercept tests and three of 30 slope tests yielded significant differences by ethnicity.

Supplementing the Current ASVAB

The results presented in this chapter point to some important considerations regarding possible supplements to the current ASVAB.

- *Spatial could add validity to AFQT and the ASVAB in general.* The Spatial and Technical scores (alone and together) provided incremental validity over AFQT for the prediction of General Technical Proficiency. When added to the regression equation, the Spatial score provided incremental validity beyond that provided by the ASVAB + Technical *R*. This finding suggests that Spatial could be a useful predictor beyond the ASVAB, not just beyond AFQT.

- *Supplements to the ASVAB could predict important criteria.* ASVAB scores were good predictors of General Technical Proficiency and had some utility for predicting Future Performance and Achievement and Effort. However, ASVAB scores were not particularly useful for predicting Physical Fitness, Teamwork, and attitudinal criteria. Thus, other predictors could provide incremental validity for predicting these criteria.

- *Supplements to the ASVAB could affect differential prediction results.* In interpreting the differential prediction results, the criterion matters. ASVAB test scores were not very useful predictors of attitudinal criteria (Tables 6.2 and 6.3), nor was the ASVAB developed for this purpose. It was developed to predict training and job performance which it does quite well. Findings of differential prediction of ASVAB scores for attitudinal variables may not be of much practical concern, except to say that other predictors designed to predict attitudinal criteria need to be considered in combination with the ASVAB. A few findings regarding prediction and differential prediction of job performance criteria merit discussion.

    - *Prediction and differential prediction of General Technical Proficiency.* As noted, ASVAB scores were good predictors of General Technical Proficiency, and this finding is consistent with our expectations for ASVAB scores based on prior research. When General Technical Proficiency was the criterion, ASVAB scores showed (a) significant overprediction of race/minority performance for three of six intercepts, no difference for the other three intercepts, and no significant slopes and (b) no significant slope or intercept differences for gender. Whether the three instances of overprediction are important depends on the organization's policies towards minorities and the current legal environment. Systemically, overprediction is undesirable because individuals who are not likely to perform well on the job will be selected. On the other hand, overprediction of race/minority performance is lenient toward the minority group because the subgroup whose predicted performance is lower than that of the referent group is treated the same as the referent group. For this reason, overprediction is often acceptable to organizations trying to recruit minorities or overcome legal challenges from minority groups.

    - *Prediction and differential prediction of Achievement and Effort, Future Expected Performance and Teamwork.* While the ASVAB scores were highly predictive of General Technical Proficiency, they were also significantly predictive, to a lesser magnitude, of Achievement and Effort, Future Expected Performance, and Teamwork (see Table 6.2). Regarding differential prediction, ASVAB scores (a) significantly underpredicted females' performance for all three of these criteria, (b) also yielded a significant gender slope difference for Achievement and Effort, and (c) tended to underpredict Hispanic performance for Teamwork and Future Expected Performance (although this finding was less salient than that of underprediction of female performance). Assuming that these criteria are important to the Army, the findings of underprediction have policy implications. By selecting on ASVAB scores alone, the Army is not selecting some females and to a lesser extent

Hispanics who are likely to work hard, be good team players, and perform well in the future Army. Since these criteria are likely to be a function of non-cognitive variables such as motivation and personality as well as cognitive ones, the underprediction might be remedied by combining the ASVAB scores with non-cognitive (i.e., personality and other) variables in the prediction equation. Clearly, it is in the Army's interest to develop, validate, and implement reliable, less fakable, measures of these non-cognitive characteristics.

CHAPTER 7: PREDICTOR SITUATIONAL JUDGMENT TEST

Gordon Waugh and Teresa Russell
HumRRO

Overview

Situational judgment tests (SJTs) have become increasingly popular in employment testing in recent years because they (a) address knowledge and skills that are difficult to measure with traditional multiple-choice test formats, (b) yield reasonably high estimated validities for predicting job performance (average r = .34 uncorrected) and incremental validity over general cognitive ability (Δr =.08 corrected) (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), and (c) typically yield small to moderate subgroup differences (Hough, Oswald, & Ployhart, 2001). SJTs provide a description of a scenario and a list of potential actions that could be taken. In some instances, the respondent reads the situation and indicates (a) which action he/she believes is *most* effective and (b) which action he/she believes is *least* effective (Weekley & Jones, 1999). Other formats have asked the respondent to indicate what he or she would be most and least likely to do in the situation (Motowidlo, Dunnette, & Carter, 1990) or to rate the effectiveness of several actions (e.g., Waugh & Russell, 2005).

Given the desirable features of SJTs, we developed a Predictor Situational Judgment Test (PSJT) for the Select21 project. Detailed information about the development of the PSJT can be found in Waugh and Russell (2005).

Instrument Description

The PSJT is a 26-item paper-and-pencil measure designed to assess the degree of good judgment in challenging situations. The situations are civilian counterparts to those typically encountered during a Soldier's first few months in the Army. Each item consists of a description of a situation followed by four actions that might be taken in that situation. The respondent rates the effectiveness of each action on a 7-point scale (see Figure 7.1).

Ineffective action.	Moderately effective action.	Very effective action.
The action is likely to lead to a bad outcome.	The action is likely to lead to a passable or mixed outcome.	The action is likely to lead to a good outcome.
—— Low ——	**———— Moderate ————**	**—— High ——**
1 2	**3 4 5**	**6 7**

Figure 7.1. PSJT response option rating scale.

The PSJT targets five dimensions: Adaptability to Changing Conditions, Relating to and Supporting Peers, Effective Self-Management, Effective Self-Directed Learning, and Teamwork. Although the PSJT items were written to reflect these dimensions, this measure was designed to yield a single total score. However, as described further in this chapter, there was a post hoc effort to develop subscores based on personality traits reflected in the PSJT response options.

Scoring

The Judgment Score

General Formula for the Judgment Score

The Soldiers responded by rating the effectiveness of each response option on a 7-point scale (where higher numbers represent greater effectiveness). We computed the judgment score for each response option using Equation 1 below.

$$Judgment\ Score_{Option\ x} = 6 - |\ Examinee Rating_{Option\ x} - keyed Effectiveness_{Option\ x}\ | \qquad (1)$$

The keyed effectiveness ratings were based on ratings by 67 subject matter experts (SMEs). The SMEs were E6 and E7 non-commissioned officers (NCOs) attending the Advanced NCO Course (ANCOC).

We subtracted the difference between the rating and keyed effectiveness values from 6 to reflect the scores, so that higher values would represent better scores. The judgment score for the entire test was the mean of the 104 option scores across the 26 scenarios.

Scoring Key Adjustments

An effectiveness rating-based scoring key has a potential disadvantage. The variability of an examinee's responses is highly correlated (in a negative direction) with the judgment scores. Because it is the average of the SMEs' effectiveness ratings, an item rarely has a keyed score of "1" or "7." There is a central tendency effect. In turn, the central tendency effect makes two relatively simple coaching strategies possible. An examinee could get a fairly good score by simply rating every option a 4 (the middle of the rating scale) or by avoiding using ratings of "1" or "7" (Cullen, Sackett, & Lievens, 2004).

In the field test (Waugh & Russell, 2005), we investigated three methods of mitigating the potential coaching effects: (a) truncating the scores, (b) stretching the key, and (c) rank ordering the scores. We found that stretching the key worked best. The algorithms for stretching the key are as follows:

For original key values above 4.0, newValue = oldValue + 0.5 * (oldValue – 4).
For original key values below 4.0, newValue = oldValue – 0.5 * (4 – oldValue).

There are advantages to using a key consisting of integers. For example, integer scores are easier to interpret, and they can be used in Item Response Theory (IRT) analyses. Therefore, after stretching the key, we rounded the new value to the nearest integer. If the new value was less than one, we rounded it up to one; if the new value was greater than 7, we rounded it down to 7.

Trait Score Development

Expert Judgments

We conducted an expert judgment exercise with 17 people from HumRRO and ARI research staff to develop the personality-based scoring scheme. The following seven KSAs were included in the exercise:

- Achievement Orientation
- Self-Reliance
- Dependability
- Affiliation/Sociability
- Agreeableness
- Social Perceptiveness
- Team Orientation

In the exercise, the experts judged the strength of the relationship (i.e., correlation) between examinees' standing on a particular trait and their effectiveness ratings for each response option. We told the experts to think of this as a correlation between the scores on a trait and the effectiveness ratings likely to be given to the response options. The experts were told to consider the traits to be perfectly measured. Each response option had five or more raters. To assess the consistency with which raters made their judgments, we computed interrater reliability estimates by form. The mean ICC(C,5) ranged from .74 to .84 for the seven traits.

We used the traitedness judgments to create a key for the PSJT. During the field test, we tried different methods of using the PSJT data to create the key. Based on several analyses, we decided to (a) allow each option to be used on no more than one trait scale and (b) have each option in a scale count equally (i.e., use unit weighting). These analyses were described in Waugh and Russell (2005).

Rasch Analyses

When we started to develop the trait scales, we had four goals. First, the scales should have at least moderate reliability. Second, the scales should be related to job performance. Third, the scales should be interpretable (i.e., reflect personality traits). Fourth, the test should be immune—or at least strongly resistant to—response distortion.

We used item response theory (IRT) analyses to develop the final trait scales. IRT has several advantages over classical item analysis (Embretson & Reise, 2000). Because of our small sample size—by IRT standards—we chose the one-parameter IRT model (i.e., the Rasch model). We used Winsteps® (2006) software to perform the analyses. Winsteps provides several diagnostic statistics that assess the dimensionality of a test and its items. These statistics helped us to develop relatively unidimensional, and thus interpretable, scales.

Using the validation data, each trait scale was fit to a Rasch (1-parameter logistic) IRT model. The data were the raw effectiveness ratings, which ranged between 1 and 7 for each option. We reversed the Soldiers' ratings (revised rating = 8– original rating) when an option was worded

such that high ratings reflected a low standing on the relevant trait. Because the ratings were not dichotomous, a polytomous Rasch model had to be used. There are two Rasch polytomous models: the *ratings* model and the *partial credit* model. The ratings model uses the same rating scale metric for every option (an "item" here refers to a PSJT response option). That is, the scale points (1-7) have the same difficulty value on the Rasch item difficulty scale for every item. In contrast, the partial credit model allows each item to have a different metric. Because of its less restrictive assumption, we used the partial credit model.

The sample size for the analyses varied from 704 to 739. Initial analyses of each trait scale showed that the rating scale points (1-7) were not equally spaced in terms of their Rasch difficulty estimates. In particular, scale points 1 through 4 were very close together. In addition, the ordering of the lower scale points (1-3) was inconsistent. That is, the ordering of the scale points conflicted with the ordering of Soldiers' scale scores. For example, in some items, Soldiers scoring lower tended to give ratings of 3 whereas Soldiers scoring higher gave ratings of 1.

Therefore, we collapsed the bottom three scale points for most of the scales. That is, rating scale points 1 through 3 were combined such that ratings of 1, 2, and 3, were changed to 4. In five scales we also collapsed scale points 4 and 5 (i.e., ratings of 4 were changed to 5). The recoded ratings, with their collapsed rating scale points, had several advantages: (a) more equal spacing of rating scale points, (b) fewer misordering of the scale points, (c) improved fit to the Rasch model, and (d) lower error variance. Table 7.1 below shows how each scale was collapsed.

Table 7.1. Rating Scale Recoding for Trait Scoring

Original Scale Points =>	1	2	3	4	5	6	7
New Scale Points: =>							
1. Achievement Orientation	4	4	4	4	5	6	7
2. Self-Reliance	4	4	4	4	5	6	7
3. Dependability	4	4	4	5	5	6	7
4. Sociability	4	4	4	5	5	6	7
5. Agreeableness	4	4	4	5	5	6	7
6. Social Perceptiveness	4	4	4	5	5	6	7
7. Team Orientation	4	4	4	5	5	6	7

After the scale points were recoded, the Rasch analyses were run. All of the trait scales had several poorly fitting items. In addition, every scale exhibited multidimensionality. Therefore, bad items were dropped in an iterative process until all of the remaining items had acceptable fit and the scales were relatively unidimensional. Specifically, the one or two worst items were dropped from a scale, and the analyses were then rerun using the revised set of items. On average, about half of the items were dropped from each scale.

Results

The PSJT was administered to 789 Soldiers. Before analyzing the data, we removed the data from 50 participants from the sample. First, 14 participants were dropped because they were observed recording their answers to the PSJT items without reading the questions. Second, 9 participants were dropped because more than 5% of their responses were missing. Third, 22 participants were dropped because their scores were very low. The frequency histogram had a

clear gap between this low-scoring group and the other participants. This cutoff score was 2.7 *SD* below the mean and was actually worse than chance-responding (which is 2.1 *SD* below the mean). Five additional Soldiers were dropped for a combination of these three reasons. The final cleaned data set contained 739 participants.

This section describes the psychometric results, estimated validities, subgroup differences, and differential prediction results for the PSJT. For a description of the methods used for each of these analyses, see Chapter 6.

Psychometric Properties

The descriptive statistics and reliability estimates for the PSJT appear in Table 7.2. The Judgment scale represents the total PSJT score (consisting of all 104 scored options). The reliability for the Judgment scale is quite high for a situational judgment test. This is due, at least in part, to the large number of response options. The McDaniel et al. (2001) meta-analyses reported reliabilities ranging from .63 to .87 with a median of .77.

Table 7.2. Descriptive Statistics for the PSJT Judgment Scale and Trait Scales

				Internal Consistency Reliability Estimates			
Scale	*k*	*M*	*SD*	Cronbach's alpha	Rasch lower-bound	Rasch upper-bound	Rasch model variance / observed variance
Judgment	104	4.66	0.33	.89	N/A	N/A	.11[a]
Achievement Orientation	13	–0.01	0.95	.85	.82	.85	.58
Self-Reliance	6	0.08	0.95	.63	.60	.66	.48
Dependability	8	0.34	1.03	.75	.71	.77	.55
Sociability	6	0.08	1.09	.72	.66	.73	.50
Agreeableness	6	0.25	1.13	.73	.68	.73	.55
Social Perceptiveness	4	0.22	1.21	.56	.55	.63	.52
Team Orientation	7	0.43	1.32	.80	.75	.80	.58

Note. *k* = number of options in the scale. For the Rasch statistics, 619 Soldiers were analyzed after dropping 20 Soldiers whose data severely misfit the Rasch model. For the other statistics, 732–738 Soldiers were analyzed after dropping Soldiers with incomplete response data.
[a]The Rasch model was not used to compute the model variance/observed variance for the Judgment score. Rather, this value (of .11) represents the proportion of variance accounted for by the first factor in a principal components analysis of the option scores.

Table 7.2 also reports Rasch reliability estimates for the trait scales. As described earlier, the seven trait scales were developed using a partial-credit Rasch model. This is a polytomous IRT model used for one-parameter logistic models. Each person has an ability value (i.e., his/her score on the construct being measured by the options in the scale), and each option has a difficulty value. The Rasch model can estimate the probability of a specific person providing a specific response to any option. Thus, every person-by-option combination has an observed response and a predicted response. The error for every response can be computed—as the difference between the observed response and the predicted response. From these errors, observed and error variances for each item and the entire scale can be computed. The Rasch

analysis can compute lower-bound and upper-bound internal consistency reliability estimates using these variances. Monte Carlo studies have shown that coefficient alpha tends to overestimate reliability and Rasch estimates tend to underestimate reliability (Linacre, 1997). Thus, the best estimate of the three estimates is likely the upper-bound Rasch estimate. Rasch reliability estimates have an advantage over coefficient alpha because they are less sample-dependent. Rasch reliabilities were not computed for the PSJT Judgment scale because the set of options in this scale did not fit the Rasch model.

The metric used for the Judgment scale differs from the metric used for the trait scales. For the Judgment scale, scores can range from 1.11 to 6.00, with random responding achieving a score of 3.81. For the trait scales, the metric is not easily interpreted. It uses a logit scale where the average level of difficulty among the items is arbitrarily given a logit score of 0. Two steps are needed to compute the logit score for a dichotomous item. First, the proportion of people getting the item right is divided by the proportion of people getting the item wrong. Second, the natural logarithm of that value is computed. Because the PSJT uses polytomous items, the odds ratio for an item is computed by dividing the number of people achieving a raw score at or above the midpoint by the number of people below the midpoint. Table 7.2 shows that most trait scales had a mean score slightly greater than one. That is because most Soldiers did well on the PSJT.

Table 7.2 also shows the proportion of modeled variance to observed variance for each of the seven trait scales. These high proportions are evidence that each trait scale was unidimensional (i.e., each trait scale was measuring one construct, although the seven different trait scales might be measuring seven different constructs). About half of the total variance in each scale was explained by the Rasch dimension. As explained above, several analyses were done to ensure that each trait scale contained only one meaningful dimension. In contrast, a factor analysis of all 104 option scores showed that the PSJT Judgment scale was multidimensional. A parallel factor analysis of the PSJT option scores suggested that the Judgment scale contains 24 factors. The first eigenvalue accounted for 28% of the common variance. Additional factor analyses were performed to extract a small number of factors. None of these solutions were interpretable.

As shown in Table 7.3, the trait scale scores were significantly correlated with each other and with the Judgment score. Interestingly, the correlation between the Judgment Score and cognitive ability as measured by AFQT (r = .22) was slightly less than what is commonly reported in the literature (i.e., r = .36; McDaniel et al., 2001). However, the meta-analysis by McDaniel et al. found a wide variation in this correlation—much wider than the variance expected due to sampling error. Only one of the trait scores, Self-Reliance, was significantly related to AFQT.

To examine the correlations among the constructs underlying the trait scales, we computed a corrected correlation matrix among the trait scales. Each correlation was corrected for unreliability in both scales. Table 7.4 shows that the underlying constructs were highly related. A principal components analysis of this corrected correlation matrix found that the first component accounted for 99.999988% of the total variance. Thus, if we assume that we have not overcorrected the correlation matrix, it seems there was only one construct underlying all of the trait scales. Thus, although each trait scale was supposed to measure a different dimension, the

scales actually were measuring the same single dimension. These results suggest that the SMEs were unable to make accurate traitedness ratings.

Table 7.3. Intercorrelations among the PSJT Judgment Scale and Trait Scales

Scale	AFQT	1	2	3	4	5	6	7	8
1. Judgment (i.e., total score))	**.22**	*.89*							
2. Achievement Orientation	.00	.42	*.85*						
3. Self-Reliance	**.12**	.31	.68	*.66*					
4. Dependability	.02	.46	.74	.59	*.77*				
5. Sociability	.02	.51	.63	.48	.59	*.73*			
6. Agreeableness	.00	.42	.76	.59	.72	.62	*.73*		
7. Social Perceptiveness	.01	.48	.56	.44	.57	.50	.55	*.63*	
8. Team Orientation	.02	.43	.74	.65	.73	.63	.74	.54	*.80*

Note. Reliability estimates for the PSJT scales are in the diagonal. n = 635 after dropping Soldiers with incomplete data, k = number of options in the scale. For the AFQT, only statistically-significant correlations are bolded ($p < .05$, two-tailed). All correlations that do not involve AFQT are statistically significant (i.e., all correlations in columns labeled 1–8).

Table 7.4. Intercorrelations among the PSJT Trait Scale Constructs

Trait Scale	6	8	2	4	3	5	7
6. Agreeableness							
8. Team Orientation	.97						
2. Achievement Orientation	.96	.90					
4. Dependability	.96	.93	.91				
3. Self-Reliance	.85	.89	.91	.83			
5. Sociability	.85	.82	.80	.79	.69		
7. Social Perceptiveness	.81	.76	.77	.82	.68	.74	

Note. n = 635 after dropping Soldiers with missing trait scores. Correlations between the traits are corrected for unreliability; thus the correlations represent the estimated correlations between the underlying constructs. Traits are listed in descending order of their average correlation with the other traits.

We also examined the construct validity of the trait scales by looking at their relationships with the Rational Biodata Inventory (RBI) scales. Before doing any analyses, we made predictions about the strengths of the correlations between the RBI scales and the PSJT scales. Our judgments were based either on (a) the degree of overlap between the constructs that the RBI and PSJT scales were trying to measure or (b) the theoretical relationship between these constructs. Table 7.5 shows the correlations between the two instruments. Considering the high intercorrelations among the PSJT trait scales, it is not surprising that this correlation matrix shows no discriminant validity. That is, all of the PSJT trait scales correlated about the same with any given RBI scale.

Table 7.5. Correlations between the PSJT and Rational Biodata Inventory (RBI)

RBI Scale	Judgment	Achieve-ment	Self-Reliance	Depend-ability	Sociability	Agree-ableness	Social Percep-tiveness	Team Orientation
Achievement	.27	**.28**	.21	.29	.28	.31	.26	.33
Army Affective Commitment	.16	.19	.13	.18	.17	.18	.12	.21
Cognitive Flexibility	.28	.26	.29	.29	.24	.28	.18	.31
Cultural Tolerance	.30	.28	.22	.27	.28	**.32**	.23	.26
Fitness Motivation	.09	.20	.14	.11	.14	.11	.11	.16
Gratitude	.30	.23	.20	.22	.25	.29	.27	.26
Hostility to Authority[S]	.35	.05	.05	.10	.06	.12	.10	.09
Internal Locus of Control	.25	.24	.19	.19	.20	.24	.20	.23
Diplomacy	.17	.25	.25	.18	**.22**	.25	.23	.26
Narcissism	-.04	.25	.16	.21	.19	.17	.20	.20
Peer Leadership	.14	**.30**	.26	.21	**.26**	.27	.19	.28
Respect for Authority	.18	.19	.14	.20	.22	.24	.14	.23
Self-Efficacy	.17	.26	**.25**	.24	.23	.22	.22	.31
Stress Tolerance	.11	.02	.01	-.04	-.03	.02	-.04	.00
Lie Scale	-.05	.15	.07	.08	.08	.12	.13	.08

Note. n = 618–645. Correlations greater than .07 are statistically significant at $p < .05$, one-tailed. Relationships that we predicted, à priori, to be strong are bold. Relationships we predicted, à priori, to be moderate are underlined. We made no predictions for the PSJT Judgment scale.

[a]The Hostility to Authority scale was reversed so that low scores represent a high level of hostility.

Criterion-Related Validity Estimates

Table 7.6 shows the zero-order correlations between the PSJT scores and the criteria. It is important to note that, unlike the validity analyses reported for the other Select21 predictors, the Achievement and Effort performance composite used in this and subsequent analyses reported in this chapter was calculated without the Criterion Situational Judgment Test (CSJT) score. Inclusion of that score would artificially inflate the validity estimates because of shared method variance with the PSJT.

The PSJT Judgment score yielded significant estimated validities for predicting all of the performance and attitudinal criteria except Physical Fitness. On the performance side, it was most closely related to Achievement and Effort and General Technical Proficiency. The corrected validity estimate for predicting General Technical Proficiency with the Judgment score was comparable to the validity estimate obtained in a prior meta-analysis (r = .34 with job performance criteria; McDaniel et al., 2001). In general, however, other performance validity estimates were lower than those obtained in the meta-analysis. Regarding attitudes, Soldiers who received high scores on the PSJT Judgment scale were relatively satisfied with the Army, fit well with the Army, and were not thinking about leaving the Army. The same pattern of estimated validities held true for the trait scales, although the levels of validity for predicting performance were generally lower for the trait scales than for the Judgment scale. There was a slight tendency for some of the trait scales (e.g., Achievement Orientation, Social Perceptiveness) to be better than the Judgment scale at predicting Physical Fitness.

Table 7.6. Criterion-Related Validity Estimates for PSJT Scores

Score	Performance Criteria					Attitudinal Criteria				
	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
Judgment Score	**.21**	**.22**	.05	**.13**	**.15**	**.28**	**.26**	**–.23**	**.12**	**.13**
Achievement Orientation	**.09**	**.17**	**.09**	**.08**	**.07**	**.24**	**.29**	**–.19**	**.22**	**.21**
Self-Reliance	**.10**	**.10**	**.06**	.03	.06	**.17**	**.21**	**–.14**	**.11**	**.17**
Dependability	.03	**.10**	.04	.05	.01	**.24**	**.29**	**–.14**	**.15**	**.24**
Sociability	.01	**.09**	.02	**.06**	.05	**.21**	**.24**	**–.18**	**.17**	**.19**
Agreeableness	.04	**.10**	.03	**.07**	.04	**.23**	**.26**	**–.16**	**.17**	**.21**
Social Perceptiveness	.05	**.11**	**.08**	.03	**.06**	**.18**	**.21**	**–.13**	**.06**	**.13**
Team Orientation	.05	**.11**	**.06**	**.07**	**.05**	**.23**	**.28**	**–.14**	**.21**	**.23**
	Corrected Validity Estimates									
Judgment Score	.33	.28	.05	.24	.26	.28	.28	–.31	.09	.12
Achievement Orientation	.10	.18	.09	.13	.09	.26	.32	–.23	.22	.22
Self-Reliance	.17	.14	.07	.07	.12	.18	.24	–.19	.10	.17
Dependability	.05	.11	.05	.09	.02	.25	.32	–.17	.15	.25
Sociability	.02	.11	.02	.11	.07	.22	.26	–.22	.17	.20
Agreeableness	.04	.11	.03	.12	.05	.25	.29	–.19	.18	.22
Social Perceptiveness	.06	.12	.08	.05	.08	.19	.24	–.16	.06	.14
Team Orientation	.07	.13	.07	.13	.07	.24	.31	–.18	.21	.25

Note. n = 648–698. Statistically significant correlations are bolded ($p < .05$, one-tailed). Corrected validity estimates have been corrected for unreliability in the criterion (first) and the indirect range restriction due to selection on the AFQT. GTP = General Technical Proficiency, AE = Achievement and Effort (without CSJT), PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Incremental Validity Estimates

As shown in Table 7.7, the PSJT Judgment score provided incremental validity for predicting all of the criteria (performance and attitudes) except one—Physical Fitness. The trait scale scores tended to add validity to the prediction of Achievement and Effort and the attitudinal criteria. Several of the trait scale scores (Achievement Orientation and Social Perceptiveness) added validity to the prediction of Physical Fitness.

We also computed the incremental validity of the trait scores after AFQT and the PSJT Judgment score had been entered into the regression equation. None of the trait scales aided prediction significantly. The largest increment in R, correcting for shrinkage, was $\Delta R = .002$, $p = .08$.

Table 7.7. Incremental Validity Estimates for PSJT Scores

Score	Performance Criteria					Attitudinal Criteria				
	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Incremental Validity Estimates									
AFQT	**.30**	**.15**	.00	**.06**	**.17**	–.01	.00	**–.12**	**–.07**	–.05
Judgment Score	**.04**	**.09**	.05	**.07**	**.04**	**.27**	**.26**	**.12**	**.08**	**.11**
Achievement Orientation	**.01**	**.07**	**.08**	**.04**	.01	**.23**	**.28**	**.11**	**.16**	**.17**
Self-Reliance	.01	**.02**	.06	.01	.00	**.16**	**.21**	**.05**	**.07**	**.14**
Dependability	.00	**.03**	.04	.02	.00	**.22**	**.28**	**.06**	**.09**	**.20**
Sociability	.00	**.02**	.02	.03	.00	**.20**	**.23**	**.10**	**.12**	**.15**
Agreeableness	.00	**.03**	.02	.03	.00	**.22**	**.26**	**.08**	**.12**	**.17**
Social Perceptiveness	.00	**.03**	**.08**	.01	.01	**.17**	**.21**	**.06**	.02	**.09**
Team Orientation	.00	**.03**	.06	.03	.01	**.21**	**.28**	**.07**	**.15**	**.19**
	Corrected Incremental Validity Estimates									
AFQT	.52	.26	.00	.16	.35	–.02	.01	–.23	–.11	–.08
Judgment Score	.02	.07	.00	.08	.02	.27	.28	.10	.05	.08
Achievement Orientation	.01	.05	.04	.03	.00	.22	.31	.09	.13	.14
Self-Reliance	.00	.01	.00	.00	.00	.15	.22	.04	.04	.11
Dependability	.00	.01	.00	.00	.00	.22	.31	.04	.06	.18
Sociability	.00	.01	.00	.01	.00	.19	.25	.08	.09	.13
Agreeableness	.00	.01	.00	.02	.00	.22	.27	.06	.09	.15
Social Perceptiveness	.00	.02	.02	.00	.00	.16	.22	.04	.00	.06
Team Orientation	.00	.02	.00	.03	.00	.21	.30	.05	.12	.17

Note. n = 648–698. Statistically significant correlations are bolded ($p < .05$, one-tailed). Uncorrected incremental estimates reflect the difference between the multiple R obtained when regressing the criterion on both the given composite and AFQT versus the R obtained when regressing the criterion only on the AFQT. Corrected incremental validity estimates have been corrected for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Cell values for the AFQT represent zero-order correlations between the AFQT and the given criterion (shown for reference). GTP = General Technical Proficiency, AE = Achievement and Effort (without CSJT), PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Finally, we computed the incremental validity of the trait scores after the PSJT Judgment score had been entered. For the Achievement and Effort composite, the Achievement Orientation and Self-Reliance trait scales had significant incremental validity. They increased validity, correcting for shrinkage, by $\Delta R = .025$. No other estimated incremental validities were significant.

Subgroup Differences

Most studies report that females score as well as or better than males on situational judgment tests (Schmitt & Chan, 2006). As Table 7.8 shows, that was certainly true for the PSJT. Female Soldiers scored significantly higher than male Soldiers on the PSJT Judgment score by about 1/2 *SD*. There was no significant difference between genders on any of the trait scales, except one. Females scored about 1/3 *SD* higher than males on Agreeableness.

Table 7.8 PSJT Scores by Gender

		Male		Female	
Score	d_{FM}	*M*	*SD*	*M*	*SD*
Judgment Score	**0.47**	4.57	0.38	4.75	0.30
Achievement Orientation	0.05	0.02	1.04	0.08	0.88
Self-Reliance	–0.04	0.08	1.00	0.04	0.85
Dependability	0.09	0.33	1.09	0.43	1.01
Sociability	0.08	0.04	1.21	0.13	0.95
Agreeableness	**0.30**	0.23	1.25	0.60	1.25
Social Perceptiveness	0.21	0.18	1.27	0.45	1.13
Team Orientation	0.04	0.44	1.42	0.49	1.21

Note. n_{Male} = 630–657. n_{Female} = 77–81. d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Prior research has reported relatively small ($d < .50$) subgroup differences with non-minority groups receiving the higher scores on situational judgment tests (Schmitt & Chan, 2006). The Select21 PSJT results were consistent with that finding, as shown in Table 7.9. No racial or ethnic subgroup difference was significant.

Table 7.9. PSJT Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
Score	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Judgment Score	–0.09	–0.04	4.60	0.37	4.57	0.40	4.61	0.36	4.59	0.38
Achievement Orientation	0.06	0.14	0.01	0.98	0.07	1.05	0.00	0.98	0.14	1.07
Self-Reliance	–0.09	–0.01	0.09	0.98	0.00	1.00	0.11	0.98	0.09	0.99
Dependability	–0.02	0.06	0.34	1.07	0.32	1.02	0.33	1.05	0.39	1.13
Sociability	–0.05	0.10	0.07	1.16	0.01	1.20	0.05	1.12	0.17	1.25
Agreeableness	0.04	0.07	0.27	1.24	0.32	1.23	0.26	1.19	0.34	1.39
Social Perceptiveness	–0.02	0.20	0.21	1.26	0.19	1.24	0.16	1.22	0.41	1.40
Team Orientation	–0.02	0.10	0.44	1.38	0.41	1.39	0.42	1.39	0.57	1.34

Note. n_{White} = 508–527. n_{Black} = 133–140. $n_{White\ Non\text{-}Hispanic}$ = 404–416. $n_{Hispanic}$ = 126–138. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites)/*SD* of Whites. None of the effect sizes are statistically significant, $p < .05$ (two-tailed).

Differential Prediction

Differential prediction results by gender, race, and ethnicity are reported in Tables 7.10, 7.11, and 7.12 respectively. In reviewing differential prediction results, there are several caveats to keep in mind. First, our sample sizes for some of the non-referent groups were smaller than what is desirable for MMR analyses. Second, we conducted a large number of analyses—480 significance tests—increasing the experiment-wide error rate. Some caution should be taken in drawing conclusions from the results.

Table 7.10. Differential Prediction Results for PSJT Scores by Gender

	Judgment					Achievement Orientation					Self-Reliance					Dependability				
	Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*	
Criterion	*b*	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F
GTP	-.09	.11	.14	**.22**	.20	-.03	.04	.14	.08	**.24**	-.01	.04	.11	**.09**	.19	-.02	.01	.04	.03	.07
AE	.13	.11	.14	**.20**	.20	**.19**	.08	.19	**.15**	**.30**	**.22**	.05	.08	**.10**	.13	**.20**	.06	.01	**.10**	.01
PF	**-.23**	.03	.22	.04	.21	-.13	.05	.20	.07	.22	-.11	.04	.08	.06	.08	-.14	.02	.19	.02	.22
TEAM	.13	.06	.11	**.11**	.14	**.18**	.03	.16	.06	**.23**	**.19**	.02	.04	.03	.05	**.19**	.03	.04	.04	.06
FXP	.08	.09	.16	**.14**	.20	.15	.03	.15	.05	.21	**.18**	.03	.04	.05	.06	**.16**	.00	.03	.01	.04
ASat	**-.34**	.22	.30	**.29**	**.34**	**-.21**	**.20**	**-.01**	**.27**	-.01	**-.19**	.14	-.04	**.19**	-.05	**-.21**	.20	.06	**.26**	.08
AFit	**-.21**	.21	.37	**.26**	**.37**	-.02	**.25**	**.01**	**.32**	.01	-.03	.17	.10	**.22**	.11	-.05	.25	.12	**.31**	.14
ACog	**.48**	-.24	-.31	**-.25**	**-.24**	**.38**	-.20	-.12	**-.21**	-.10	**.35**	-.13	-.13	**-.15**	-.11	**.35**	-.14	-.11	**-.15**	-.10
CInt	-.08	.13	.20	**.12**	.14	.04	.26	.00	**.24**	.00	.04	.14	-.06	**.13**	-.05	-.03	.19	-.01	**.16**	-.01
FAA	**-.36**	.14	.15	**.15**	.13	**-.25**	**.22**	**-.11**	**.24**	-.11	**-.25**	**.18**	**-.19**	**.20**	-.20	**-.28**	**.25**	**-.02**	**.27**	-.02

	Sociability					Agreeableness					Social Perceptiveness					Team Orientation				
	Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*		Gen	PSJT *b*		*r by Gen*	
Criterion	*b*	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F
GTP	-.03	**-.01**	**.15**	-.01	**.24**	-.07	**.00**	**.16**	.01	**.31**	-.03	.03	-.01	.06	-.02	-.03	**.01**	**.16**	.02	**.28**
AE	**.19**	.04	.18	.07	**.27**	**.15**	**.03**	**.17**	.06	**.31**	**.19**	.05	.04	**.10**	.06	**.20**	.05	.15	**.09**	**.23**
PF	-.14	.01	.11	.01	.11	-.15	.01	.11	.02	.13	-.15	.06	.08	**.08**	.09	-.14	**.03**	**.26**	.04	**.28**
TEAM	**.18**	.02	.16	.04	.21	**.15**	.02	.16	.04	**.25**	**.16**	.01	.06	.01	.08	**.18**	.03	.14	.05	.21
FXP	.15	.01	.19	.02	**.25**	.12	.00	.15	.01	**.24**	.15	.04	-.02	.06	-.03	**.16**	.02	.13	.03	.19
ASat	**-.23**	.16	.15	**.22**	.17	**-.19**	**.22**	**-.03**	**.28**	-.05	**-.19**	**.16**	**-.07**	**.21**	-.09	**-.19**	**.19**	**-.02**	**.25**	-.02
AFit	-.07	.19	.22	**.24**	.22	-.05	**.24**	**.03**	**.29**	.04	-.03	.18	.03	**.23**	.04	-.02	.24	.05	**.30**	.06
ACog	**.38**	-.17	-.30	**-.18**	**-.23**	**.38**	-.19	-.04	**-.20**	-.04	**.38**	-.14	-.02	**-.15**	-.02	**.36**	-.14	-.13	**-.15**	-.11
CInt	-.01	.19	.23	**.17**	.16	.00	.22	-.01	**.20**	-.01	.03	**.09**	**-.20**	**.08**	-.16	.02	.26	.00	**.23**	.00
FAA	**-.29**	.19	.06	**.21**	.05	**-.23**	**.25**	**-.17**	**.27**	-.20	**-.27**	.14	-.04	**.16**	-.05	**-.24**	**.25**	**-.14**	**.27**	-.14

Note. $n_{Regression}$ = 647–697. n_{Male} = 572–624. n_{Female} = 69–79. Gen *b* = Unstandardized regression weight for gender (0 = male, 1 = female).
PSJT *b* = Unstandardized regression weight for the given PSJT scale for males and females. *r* by Gen = Correlation between the given PSJT scale and the given criterion for each gender. Regression weights for males and females are bolded if the PSJT-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort (without CSJT), PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 7.11. Differential Prediction Results for PSJT Scores by Race

	Judgment					Achievement Orientation					Self-Reliance					Dependability				
	Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race	
Criterion	*b*	W	B	W	B	*b*	W	B	W	B	*b*	W	B	W	B	*b*	W	B	W	B
GTP	**-.25**	.11	.10	**.20**	**.25**	**-.26**	.05	.09	.09	**.20**	**-.25**	.05	.07	**.10**	.17	**-.26**	.00	.09	.00	**.18**
AE	**-.15**	.14	.11	**.25**	**.19**	**-.17**	.09	.15	**.17**	**.27**	**-.16**	.08	.04	**.14**	.07	**-.16**	.06	.06	**.11**	.10
PF	-.07	.06	-.03	.07	-.04	-.08	**.11**	**-.04**	**.13**	-.05	-.07	**.10**	**-.10**	**.12**	-.14	-.08	**.07**	**-.13**	**.09**	-.16
TEAM	-.03	.09	.08	**.16**	.14	-.05	.02	.09	.04	.15	-.04	.01	.04	.01	.07	-.03	.02	.11	.03	.17
FXP	**-.17**	.13	.03	**.19**	.06	**-.18**	.05	.05	.07	.08	**-.19**	.06	.01	**.09**	.02	**-.17**	.01	.01	.02	.02
ASat	-.04	.23	.21	**.29**	**.28**	-.04	.20	.18	**.25**	**.24**	-.06	.13	.10	**.17**	.13	-.07	.19	.14	**.25**	.16
AFit	-.12	.23	.17	**.28**	**.23**	-.14	.25	.17	**.30**	**.23**	-.14	.17	.10	**.21**	.14	**-.17**	.25	.13	**.30**	.15
ACog	**.37**	-.24	-.14	**-.24**	-.16	**.39**	-.22	-.10	**-.22**	-.11	**.40**	-.14	-.11	**-.14**	-.12	**.42**	-.15	-.02	**-.15**	-.02
CInt	.07	**.20**	**-.01**	**.17**	-.01	.03	.28	.14	**.24**	.13	.05	.14	.01	**.13**	.01	.08	.18	.03	**.15**	.03
FAA	-.15	.13	.11	**.14**	.12	-.17	.17	.24	**.18**	**.27**	-.14	.11	.22	**.12**	**.24**	-.17	.21	.19	**.23**	**.20**

	Sociability					Agreeableness					Social Perceptiveness					Team Orientation				
	Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race		Race	PSJT *b*		*r* by Race	
Criterion	*b*	W	B	W	B	*b*	W	B	W	B	*b*	W	B	W	B	*b*	W	B	W	B
GTP	**-.26**	-.01	.05	-.03	.11	**-.27**	.00	.10	.00	**.22**	**-.26**	.04	.01	.07	.03	**-.26**	.02	.06	.04	.15
AE	**-.15**	.04	.08	.08	.14	**-.17**	.06	.08	**.11**	.14	**-.16**	.09	.00	**.16**	-.01	**-.15**	.07	.09	**.12**	.16
PF	-.07	.04	-.09	.05	-.12	-.06	.05	-.10	.06	-.13	-.06	.08	.02	**.11**	.03	-.09	**.11**	**-.07**	**.14**	-.10
TEAM	-.03	.03	.07	.05	.12	-.04	.05	.08	.08	.12	-.04	.04	-.03	.07	-.05	-.03	.04	.07	.07	.11
FXP	**-.16**	.02	.07	.04	.13	**-.18**	.03	.02	.05	.04	**-.17**	.07	-.04	**.11**	-.06	**-.17**	.05	.05	.07	.09
ASat	-.05	.17	.20	**.22**	**.26**	-.06	.19	.17	**.25**	**.21**	-.05	.15	.15	**.19**	**.19**	-.04	.17	.24	**.22**	**.31**
AFit	-.13	.20	.21	**.25**	**.28**	**-.15**	.24	.12	**.29**	.15	-.15	.15	.19	**.20**	**.25**	-.13	.23	.18	**.28**	**.23**
ACog	**.38**	-.21	-.09	**-.22**	-.09	**.41**	-.19	-.07	**-.19**	-.07	**.42**	-.12	-.13	**-.12**	-.14	**.38**	-.15	-.07	**-.16**	-.08
CInt	.07	.25	.11	**.21**	.10	.07	**.25**	**-.02**	**.21**	-.02	.04	.08	-.03	.07	-.03	.05	.27	.13	**.23**	.11
FAA	-.15	.22	.08	**.23**	.09	**-.18**	.20	.23	**.22**	**.24**	-.16	.11	.17	**.12**	**.18**	-.15	.18	.31	**.20**	**.33**

Note. $n_{Regression}$ = 591–630. n_{White} = 469–500. n_{Black} = 121–130. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black).
PSJT *b* = Unstandardized regression weight for the given PSJT scale for Whites and Blacks. *r* by Race = Correlation between the given PSJT scale and the given criterion for each race. Regression weights for Whites and Blacks are bolded if the PSJT-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort (without CSJT), PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 7.12. Differential Prediction Results for PSJT Scores by Ethnic Group

	Judgment					Achievement Orientation					Self-Reliance					Dependability				
	Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*	
Criterion	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H
GTP	-.06	.12	.07	**.22**	.15	-.08	.07	-.02	**.13**	-.06	-.06	**.10**	**-.06**	**.17**	-.14	-.06	.03	-.06	.06	-.14
AE	**.15**	.15	.06	**.26**	.12	**.13**	.10	.06	**.17**	.12	**.15**	**.12**	**-.03**	**.20**	-.05	**.15**	.08	.01	**.13**	.01
PF	.08	.06	.07	.07	.10	.09	.12	.04	**.15**	.06	.08	.12	.00	**.15**	.00	.10	.08	.05	.10	.08
TEAM	**.19**	.09	.07	**.15**	.15	**.19**	.04	.01	.06	.02	**.19**	**.06**	**-.11**	**.10**	**-.22**	**.19**	.03	-.02	.05	-.04
FXP	.08	.15	.04	**.21**	.07	.06	.07	-.01	**.09**	-.02	.07	**.12**	**-.07**	**.16**	-.12	.07	.03	-.03	.05	-.06
ASat	.11	**.25**	**.07**	**.31**	.09	.01	.19	.22	**.23**	**.31**	.10	.13	.14	**.17**	**.18**	.07	.18	.20	**.22**	**.27**
AFit	.11	.22	.20	**.26**	**.25**	.04	.23	.28	**.27**	**.39**	.11	.14	.24	**.18**	**.32**	.06	.22	.32	**.26**	**.42**
ACog	.00	**-.28**	**-.08**	**-.28**	-.08	.07	-.23	-.19	**-.22**	**-.22**	.02	-.13	-.22	**-.14**	**-.23**	.05	-.16	-.12	**-.16**	-.14
CInt	.01	.20	.05	**.16**	.05	-.11	.26	.31	**.21**	**.32**	-.02	.12	.23	**.11**	**.22**	-.04	.14	.26	**.11**	**.25**
FAA	.17	.13	.12	**.14**	.14	.13	.18	.17	**.18**	**.22**	.18	.07	.25	.08	**.31**	.12	.20	.26	**.21**	**.32**

	Sociability					Agreeableness					Social Perceptiveness					Team Orientation				
	Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*		Eth	PSJT *b*		*r by Eth*	
Criterion	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H
GTP	-.06	.01	-.06	.01	-.13	-.07	**.03**	**-.07**	.06	**-.18**	-.06	.06	-.03	**.10**	-.07	-.06	.05	-.06	.08	-.12
AE	**.14**	.07	-.01	**.12**	-.03	**.15**	**.09**	**-.02**	**.16**	-.05	**.14**	.10	.01	**.18**	.02	**.16**	.08	.01	**.15**	.02
PF	.08	.04	.05	.05	.07	.07	.05	.02	.06	.03	.08	.08	.06	**.10**	.09	.07	.10	.10	**.13**	.13
TEAM	**.19**	.03	.01	.05	.03	**.18**	.07	-.01	**.12**	-.03	**.19**	.05	-.01	.08	-.02	**.19**	.06	-.01	.10	-.02
FXP	.07	.03	-.01	.05	-.01	.06	.05	-.01	.07	-.02	.08	.09	-.03	**.13**	-.05	.06	.07	-.03	.10	-.06
ASat	.09	.18	.10	**.23**	.14	.09	**.23**	**.09**	**.29**	.12	.09	.17	.10	**.22**	.16	.12	.17	.17	**.22**	**.23**
AFit	.09	.22	.13	**.26**	**.18**	.09	.23	.22	**.27**	**.31**	.06	.14	.22	**.17**	**.32**	.10	.21	.29	**.25**	**.38**
ACog	.04	-.24	-.15	**-.24**	-.17	.03	-.19	-.17	**-.19**	**-.20**	.05	-.14	-.07	**-.14**	-.08	.04	-.15	-.21	**-.15**	**-.23**
CInt	-.02	.25	.19	**.20**	**.18**	-.04	.25	.24	**.20**	**.25**	-.03	.08	.09	.07	.10	-.03	.24	.34	**.20**	**.31**
FAA	.14	.21	.25	**.21**	**.31**	.12	.20	.21	**.21**	**.28**	.14	.10	.13	**.11**	.18	.15	.17	.19	**.19**	**.22**

Note. $n_{\text{Regression}}$ = 486–524. $n_{\text{White non-Hispanic}}$ = 374–392. n_{Hispanic} = 109–132. Eth *b* = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). PSJT *b* = Unstandardized regression weight for the given PSJT scale for White non-Hispanics and Hispanics. *r* by Eth = Correlation between the given PSJT scale and the given criterion for each ethnic group. Regression weights for White non-Hispanics and Hispanics are bolded if the PSJT-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).GTP = General Technical Proficiency, AE = Achievement and Effort (without CSJT), PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Judgment Score

The Judgment score merits special attention since it is more likely than the personality scores to be used in the future. Only three out of 30 slope tests for the Judgment score were significant. About one-third of the 30 intercept tests for regressing either performance or attitudinal criteria on the Judgment score were significant. The intercept tests suggested that the Judgment scale:

- overpredicted female Soldiers' Physical Fitness scores;
- overpredicted Black Soldiers' performance on General Technical Proficiency, Achievement and Effort, and Future Performance;
- underpredicted Hispanic Soldiers' performance on Achievement and Effort and on Teamwork;
- overpredicted female Soldiers' satisfaction, fit, and future Army affect and underpredict their attrition cognitions; and
- underpredicted Black Soldiers' attrition cognitions.

Trait Scores

Notably, nearly half of the slope tests for regressing performance criteria on the trait scores were significant, making interpretation of results for the trait scores difficult. Intercept results for the trait scores were very similar to results for the Judgment score with one notable difference. All seven trait scores underpredicted females' performance on two of the five performance criteria (i.e., Achievement and Effort and Teamwork), and three of the trait scores also underpredicted Future Expected Performance.

Summary

Judgment Score

The PSJT Judgment score yielded significant estimated validities for predicting all of the performance and attitudinal criteria except Physical Fitness. On the performance side, it was most closely related to Achievement and Effort (computed without CSJT) and General Technical Proficiency. Regarding attitudes, Soldiers who received high PSJT Judgment scores were relatively satisfied with the Army, fit well with the Army, and were not thinking about leaving the Army. This score provided incremental validity over AFQT for predicting all of the criteria (performance and attitudes) except one—Physical Fitness. Females received Judgment scores that were .47 *SD* higher than males' scores, and there were no significant race or ethnic differences. The analyses of differential prediction suggested that the Judgment score yielded few slope differences for predicting the criteria. About one-third of the intercept differences were statistically significant.

Trait Scores

Overall, the results for the trait scales were disappointing. The analyses appeared to show that the seven trait scales measure the same construct—and we do not know what that construct is. They yielded lower estimated validities than the Judgment score for predicting the criteria and

added little or no incremental validity beyond AFQT. Thus, the usefulness of the trait scales is doubtful. Perhaps the traitedness ratings were flawed or the wrong set of traits was chosen. Alternatively, it might just be that the set of constructs underlying Soldiers' option ratings are too complex to adequately measure.

Issues Regarding Operational Use

The PSJT is likely to be useful and easy to administer operationally. It is automated and relatively simple to score. As mentioned previously, the Judgment score yielded significant validity for predicting most criteria. The trait scores, however, are not likely to be useful.

The SJT format does create a challenge for selection testing programs. Organizations, particularly those who test and retest large numbers of applicants, use alternate forms to increase form security and decrease retest effects. Alternate forms for situational judgment tests cannot be constructed in the same manner as traditional tests. A domain sampling method is usually used to develop alternate forms for traditional tests. In this approach, item authors target a specific construct or content domain for each item. For situational judgment tests, however, little is known about the test's underlying constructs or content domains. Recent research suggests that alternate forms for situational judgment tests must be cloned at the item level. Lievens and Sackett (2006) compared three different approaches to constructing alternate forms: domain sampling, incident cloning, and item cloning. In the incident sampling approach, a new specific situation was written based on the same general critical incident. In the item cloning approach, only cosmetic changes in wording were made. The alternate form correlations for the three methods were .22, .41, and .57 for the domain sampling, critical incident cloning, and item cloning methods, respectively. Thus, the alternate reliability estimate was unacceptable if substantive changes were made to the items. When test forms are this similar, it is possible that people who retest on the alternate form might score higher than they would retesting on substantively different alternate forms. Lievens and Sackett found that scores increased only slightly on the second (taken a year later) item-cloned alternate form (d = 27), and scores on the second incident-cloned alternate form actually increased much more (d = .67). Thus, the item-cloning strategy appears to be a strategy worth pursuing.

CHAPTER 8: WORK SUITABILITY INVENTORY

Rodney A. McCloy and Dan J. Putka
HumRRO

Overview

The primary stumbling block for personality measures has been their tendency to predict performance well in research settings but to have reduced validity in operational settings (Knapp, Waters, & Heggestad, 2002). One reason frequently given for this phenomenon is response distortion—the capacity and tendency of respondents to answer in a dishonest fashion, usually with an eye toward presenting themselves as they believe the organization would like them to appear. Recent efforts to combat response distortion have focused on innovative response formats, such as multidimensional forced-choice measures (Jackson, Wrobleski, & Ashton, 2000; Sisson, 1948; White & Young, 1998; Wright & Miederhoff, 1999). In this chapter, we present concurrent validation (CV) results for the Work Suitability Inventory (WSI), a personality measure that also adopts a unique format and corresponding set of scoring procedures aimed at reducing the deleterious effects of response distortion on the validity of personality assessments.

Instrument Description

The WSI incorporates a computerized card-sorting task in which respondents sort 16 statements describing different types of work requirements. To give a sense of the types of statements respondents encounter, the following two statements appear on the WSI:

- Work that requires…showing a cooperative and friendly attitude towards others I dislike or disagree with.
- Work that requires…being open to change (positive or negative) and a lot of variety.

Each statement appears on its own rectangular block, or "card." Respondents rank the 16 statements in terms of how well they think they would perform each type of work described—the highest ranked statement should describe work that respondents think they would perform best, and the lowest ranked statement should describe work that respondents think they would perform least well (see McCloy & Putka, 2005, for more detailed information regarding development of the WSI). Each of the 16 statements is tied to a personality trait or "work style" (Borman, Kubisiak, & Schneider, 1999).

The most important feature of the WSI regards its scoring: The Army can score the WSI differently for each outcome (e.g., job performance, attrition, person-Army fit) it predicts. Unlike conventional tests having correct answers or a single set of keyed answers, no single ordering of the 16 cards will result in a highest WSI score for all outcomes to be predicted—a ranking that yields a high score on one outcome may well yield a low score on other outcomes. Therefore, applicants' attempts to rank the statements the way they think the Army would like them to (rather than ranking them in the way that best describes them) will be counterproductive unless the applicants know the scoring algorithm for the outcome of interest (e.g., to get assigned to a particular job). Of course, this feature alone cannot prevent respondents from responding in a

dishonest fashion, but we believe it will reduce the frequency of prevarication. As such, the WSI is resistant to tampering and response distortion.

Method

Sample

A total of 783 Soldiers completed the WSI during the concurrent validation data collections (Wave 1 = 606, Wave 2 = 177). We cleaned these data using three primary screens: (a) Soldiers deemed to take too little time to complete the WSI (i.e., less than 140 seconds), (b) Soldiers whom the data log reported as malingering during the data collection session. and (c) Soldiers displaying an unlikely response pattern. Regarding these response patterns, we targeted one in which the first four or last four cards were consecutive—that is, the top-ranked cards were A, B, C, and D and/or the lowest ranked cards were M, N, O, and P. This pattern search subsumes those who simply ranked cards A through P as 1 through 16 respectively (i.e., each card was sorted into the nearest box). A breakdown of the number of Soldiers eliminated based on each screen is shown in Table 8.1. The final analysis sample, therefore, comprised 682 Soldiers (Wave 1 = 523, Wave 2 = 159).

Table 8.1. Summary of the Total and Cleaned Concurrent Validation Samples for the WSI

	Wave 1	Wave 2	Total
Total CV Sample	606	177	783
Total Deletions	83	28	101
Time Deletion	68	14	82
Pattern Deletion	28[a]	0	28
Problem Log Deletion	1	4	5
Cleaned CV Sample	523	159	682

[a]Fourteen of these 28 Soldiers were also flagged for deletion on the "too little time" screen. Therefore, the total number of deletions (101) is 14 less than the total number of Soldiers identified by each screen across Waves 1 and 2.

Validation Strategy

The WSI employs an empirical keying procedure, identifying a best composite for each criterion variable. As described in Chapter 5 of this report, 10 criterion composites—five assessing job performance and five assessing Soldier attitudes—were selected for use in the criterion-related validity analyses. Maximum insurance against response distortion would occur if each criterion could be linked to the placement of unique combinations of the 16 WSI cards. Therefore, we hoped to attain a reasonable degree of differential validity for the various criterion composites.

Although many scores could be calculated for the WSI, the CV analysis investigated only two types. The first type is a "full score" for each of the 16 dimensions. This full score is simply the rank of the dimension subtracted from 17. Thus, if a respondent ranked dimension C (Attention to Detail) third out of the 16 cards, that card would receive a full score of 17 – 3 = 14. The second was an optimal, empirically keyed composite of "dyad scores." A dyad score is a dichotomous variable that indicates whether a given dimension was ranked higher than another dimension.

Given 16 dimensions, there are (16*15) / 2 = 120 unique pairs of dimensions. Because dimension 1 could be ranked higher or lower than dimension 2, each pair of WSI statements requires two dyad scores. Thus, a total of 2*120 = 240 dyad scores were calculated. The most predictive set of dyads was identified for each of the 10 criteria examined in the CV analysis—hence, the CV analysis investigated the predictive validity of 10 empirical dyad composite (EDC) scores. To reduce some of the capitalization on chance that was surely occurring, we applied unit weights, rather than the optimal weights obtained from the regression analyses, to the dyad scores. Hence, the dyads were summed, and those sums are the predictor scores for which validity estimates were obtained.

Cross-Validation

The unabashedly empirical approach to developing EDCs demands that they be cross-validated so that any differential validity across predictor scores can be attributed to true variation in predictive strength and not to vagaries of the development sample. Although the EDCs were unit-weighted, their content was based on an optimal empirical procedure (stepwise regression). As a result, we would expect the criterion-related validity of these composites to shrink upon application—both in another sample and (to a lesser extent) in the population.

Given that the construction of all the "weighted" composites was at least partially based on the data, it would be desirable to have adjusted validity estimates that account for the shrinkage that is likely upon cross-validation. Under typical circumstances, the preferred approach would be to apply a shrinkage formula to the criterion-related validity estimate obtained in the full sample (e.g., Cattin, 1980). However, there were several factors which made application of such formulae hazardous in this case: (a) the multiple steps involved in the process of forming the regression weighted composites noted above, and (b) the partial dependence of the subjectively weighted and unit composites on the regression results. In light of the questionable nature of formula-based shrinkage corrections for composites such as these, we adopted an alternative strategy for cross-validation.

As described in Chapter 2, CV data were collected in two waves. Data collected in the first wave were used as a calibration sample in which we established the content of the empirical dyad composites described above. Data collected in the second wave were used as a cross-validation sample in which we took the models developed in Wave 1 and applied them to the Wave 2 data. This approach allowed us to calculate criterion-related validity estimates in Wave 1, and cross-validated criterion-related validity estimates in Wave 2. Finally, we used the total CV sample to revisit the content of all EDCs based on all of the data available. Revisiting the content and weighting of these composites based on the full sample allowed us to obtain the most stable estimates possible for the EDCs. Although composites based on the full CV sample are of ultimate focus in this and subsequent chapters that conduct cross-instrument analyses, comparison of Wave 1 and Wave 2 validity estimates gives the reader an idea of how stable the full sample results might be in subsequent independent samples.[22]

[22] We want to emphasize that comparison of Wave 1 and Wave 2 results will only provide a rough estimate of how well the full sample composites would be expected to cross-validate. First, all else being equal, the full CV sample results should be more stable than those based on Wave 1 (simply due to a larger sample size). Also, given that the content of the EDCs based the final EDCs on results from the full CV sample, it is possible that the final EDCs will differ from the Wave 1 EDCs, even for those composites targeting the same criterion.

Results

WSI Full Scores

Table 8.2 shows descriptive statistics for the WSI full scores. The table indicates that two of the 16 statements (Stress Tolerance and Persistence) received relatively low rankings from the majority of Soldiers. Although Achievement and Effort received the highest mean ranking, several other cards also received high ranks from the Soldiers (Leadership Orientation, Independence, Attention to Detail, Innovation), with means that fall within one point of the mean for the top-ranked statement.

Table 8.2. Descriptive Statistics for WSI Full Scores

WSI Dimension	Minimum	Maximum	*M*	*SD*
A: Achievement and Effort	1	16	10.46	4.59
B: Adaptability/Flexibility	1	16	8.84	4.37
C: Attention to Detail	1	16	9.92	4.33
D: Concern for Others	1	16	7.22	4.91
E: Cooperation	1	16	7.56	4.43
F: Dependability	1	16	8.76	4.18
G: Energy	1	16	9.21	4.36
H: Independence	1	16	9.97	4.86
I: Initiative	1	16	7.42	3.75
J: Innovation	1	16	9.61	4.39
K: Leadership Orientation	1	16	10.28	4.22
L: Persistence	1	16	6.61	4.06
M: Self-Control	1	16	7.77	4.28
N: Social Orientation	1	16	8.47	4.62
O: Stress Tolerance	1	16	5.88	4.36
P: Cultural Tolerance	1	16	8.01	4.84

Note. $n = 682$.

Correlations among the WSI full scores appear in Table 8.3. For the most part, these correlations are quite low and negative, the latter characteristic stemming from the ipsative nature of the scores.

Gender Differences

Men and women frequently score differently on personality traits (e.g., Costa, Terracciano, & McCrae, 2001; Linz & Semykina, in press; Lynn & Martin, 1997; Srivastava, John, Gosling, & Potter, 2003). We therefore examined the degree to which male and female Soldiers ranked the WSI statements differently. One means of doing so involved calculating effect sizes (i.e., *d* statistics) for each WSI full score by subtracting the mean rank for men from the mean rank for women and dividing by the standard deviation for men (i.e., the referent group; women are the focus group). The other approach entailed ranking the full scores by their means within gender and comparing the ranks. This second approach provided a different view of the responses provided by male and female Soldiers in the CV sample.

Table 8.4 contains the mean ranks given to each WSI full score by female and male Soldiers. In keeping with the two approaches used to examine the ranking data, the table presents

the results in two ways—first by order of the full score and then by order of the mean ranks. The *d* statistics associated with the first ordering indicate that women had higher mean ranks for the WSI full scores Concern for Others, Cooperation, and Cultural Tolerance; men had higher ranks for Energy, Persistence, Self-Control, and Stress Tolerance. The second view shows more clearly that there was a bit more variability in the ranks of the female Soldiers (means ranging from 6.31 to 12.73) than the ranks of the male Soldiers (means ranging from 6.57 to 10.89).

Table 8.3. Intercorrelations among WSI Full Scores

	WSI Dimension														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
B	**.11**														
C	**.20**	.02													
D	**-.16**	**.09**	**-.10**												
E	**-.11**	**.08**	-.01	**.34**											
F	**.10**	**-.11**	**.17**	**-.07**	**-.07**										
G	.01	**-.08**	**-.08**	**-.24**	**-.13**	.05									
H	-.05	**-.11**	-.07	**-.18**	**-.19**	.04	.03								
I	**-.10**	-.06	-.01	**-.15**	**-.17**	-.02	-.02	-.06							
J	**-.14**	**-.12**	**-.21**	.00	**-.18**	**-.23**	**-.14**	**.09**	-.01						
K	**-.10**	**-.27**	**-.14**	**-.24**	**-.22**	**-.08**	.03	.01	**.08**	**.10**					
L	-.06	**-.15**	-.03	**-.18**	**-.19**	**-.09**	**-.10**	.03	-.03	-.04	.02				
M	**-.14**	**-.17**	**-.17**	**-.19**	**-.12**	**-.14**	-.07	**-.09**	-.05	-.03	.01	**.08**			
N	**-.22**	-.05	**-.21**	.05	-.02	**-.19**	**-.11**	**-.29**	**-.15**	.02	-.02	**-.08**	.03		
O	**-.09**	**-.15**	**-.08**	**-.32**	**-.16**	-.02	**.09**	-.07	.01	**-.20**	-.01	**.08**	**.14**	-.02	
P	**-.27**	-.06	**-.22**	**.18**	.07	**-.27**	**-.18**	**-.14**	**-.08**	.05	**-.09**	**-.16**	-.04	**.18**	**-.13**

Note. $n = 682$. Statistically significant correlations are bolded ($p < .05$, two-tailed). A = Achievement and Effort, B = Adaptability/Flexibility, C = Attention to Detail, D = Concern for Others, E = Cooperation, F = Dependability, G = Energy, H = Independence, I = Initiative, J = Innovation, K = Leadership Orientation, L = Persistence, M = Self-Control, N = Social Orientation, O = Stress Tolerance, P = Cultural Tolerance.

Table 8.4. Mean Ranks by Gender for the WSI Full Scores

Females		Males	$d_{F\text{-}M}$	Females		Males	
6.31	Achievement and Effort	6.57	-0.06	Achievement and Effort	6.31	Achievement and Effort	6.57
7.63	Adaptability/Flexibility	8.25	-0.14	Cultural Tolerance	6.66	Leadership Orientation	6.61
7.17	Attention to Detail	7.10	0.02	Concern for Others	7.01	Independence	6.97
7.01	Concern for Others	10.10	-0.64	Attention to Detail	7.17	Attention to Detail	7.10
8.10	Cooperation	9.55	-0.33	Leadership Orientation	7.24	Innovation	7.34
8.04	Dependability	8.29	-0.06	Independence	7.56	Energy	7.65
9.29	Energy	7.65	0.38	Adaptability/Flexibility	7.63	Adaptability/Flexibility	8.25
7.56	Independence	6.97	0.12	Innovation	7.69	Dependability	8.29
9.71	Initiative	9.57	0.04	Social Orientation	8.00	Social Orientation	8.59
7.69	Innovation	7.34	0.08	Dependability	8.04	Self-Control	9.02
7.24	Leadership Orientation	6.61	0.15	Cooperation	8.10	Cultural Tolerance	9.25
11.86	Persistence	10.23	0.40	Energy	9.29	Cooperation	9.55
11.00	Self-Control	9.02	0.46	Initiative	9.71	Initiative	9.57
8.00	Social Orientation	8.59	-0.13	Self-Control	11.00	Concern for Others	10.10
12.73	Stress Tolerance	10.89	0.41	Persistence	11.86	Persistence	10.23
6.66	Cultural Tolerance	9.25	-0.54	Stress Tolerance	12.73	Stress Tolerance	10.89

Note. $n_{females} = 70$, $n_{males} = 604$. $d_{F\text{-}M}$ = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Negative effect sizes indicate that females ranked the full score higher than did males.

Table 8.5 presents a slightly different view of the mean ranks. Specifically, it assigns (within gender) a rank from 1 to 16 for each of the mean ranks. The difference between these ranks highlights those WSI statements with the most discrepant rank orders for men and women. The table shows that women ranked Concern for Others (score D) and Cultural Tolerance (score P) much higher than men did (ranks of 3 and 2 for women, respectively, versus ranks of 14 and 11 for men). Men ranked several of the dimensions higher than women, although the difference was greatest for Energy (score G) and Self-Control (score M). Note, however, that men and women ranked 7 of the 16 scores the same. Also of note is that the relatively high *d* statistics for Persistence (score L) and Stress Tolerance (score O) showing that men had higher mean ranks for these statements (see Table 8.4) did *not* translate into differential overall ranks—men and women both ranked these as the penultimate and lowest statements.

Table 8.5. Rank Orders of the Mean Ranks by Gender for the WSI Full Scores

WSI Full Score	Females	Males	Difference in Ranks (F – M)
Achievement and Effort	1	1	0
Adaptability/Flexibility	7	7	0
Attention to Detail	4	4	0
Concern for Others	3	14	-11
Cooperation	11	12	-1
Dependability	10	8	2
Energy	12	6	6
Independence	6	3	3
Initiative	13	13	0
Innovation	8	5	3
Leadership Orientation	5	2	3
Persistence	15	15	0
Self-Control	14	10	4
Social Orientation	9	9	0
Stress Tolerance	16	16	0
Cultural Tolerance	2	11	-9

Note. $n_{females} = 70$, $n_{males} = 604$.

WSI Empirical Dyad Composites

In addition to the full scores, we also calculated optimal composites for predicting each of the 10 criterion composites. These WSI composites consist of dyad scores. As mentioned previously, a dyad score is a dichotomous variable indicating whether a given WSI dimension was ranked higher than another dimension. The dyads were selected using a purely empirical procedure, hence the term "empirical dyad composite" (EDC) for the resulting scores. Specifically, the procedure for developing the EDCs was as follows:

- Calculate the zero-order correlation between each criterion and each dyad score.
- Select candidate dyads that exceed some minimum correlation with the criterion.[23]
- Enter the candidate dyads into a backward elimination stepwise regression program.[24]

[23] The cutoff was ±.095 for all criteria except Teamwork and Expected Future Performance, where the cutoff was lowered to ±.065 so that a reasonable number of dyads could be retained for further consideration.

[24] The *p* value for entry was set to .05 and for elimination once entered to .10. Hence, we were more stringent about dyads entering the EDC than about removing them once entered.

- Calculate the EDC as the simple sum of the dyads retained by the stepwise regression procedure, thus applying unit weights to the dyads rather than the optimal stepwise regression weights.

Descriptive Statistics

Tables 8.6 and 8.7 contain descriptive statistics and intercorrelations for the 10 WSI EDCs, respectively. Table 8.6 shows that the number of dyads within each composite ranges from a minimum of two (the Teamwork EDC) to a maximum of eight (the EDC for Satisfaction with the Army) with the modal number of dyads being five. Table 8.7 shows that the dyads did evidence some intercorrelation, especially among the EDCs for the five attitudinal composite criteria. This finding is not surprising given the moderate to high correlations observed between the attitudinal criteria (see Chapter 3).

Table 8.6. Descriptive Statistics for the WSI Empirical Dyad Composites

Empirical Dyad Composite	Minimum	Maximum	*M*	*SD*
Predictor for Future Expected Performance	0	4	2.29	0.90
Predictor for General Technical Proficiency	0	5	2.59	1.21
Predictor for Achievement and Effort	0	5	2.90	1.17
Predictor for Physical Fitness	0	4	2.10	0.93
Predictor for Teamwork	0	2	0.92	0.70
Predictor for Satisfaction with the Army	0	8	4.06	1.64
Predictor for Perceived Army Fit	0	7	3.47	1.45
Predictor for Attrition Cognitions	0	5	2.52	1.23
Predictor for Career Intentions	0	5	2.30	1.22
Predictor for Future Army Affect	0	7	3.75	1.49

Note. n = 682.

Table 8.7. Intercorrelations of the WSI Empirical Dyad Composites

	Empirical Dyad Composite								
	FXP	GTP	AE	PF	TEAM	ASat	AFit	ACog	CInt
General Technical Proficiency (GTP)	**.50**								
Achievement and Effort (AE)	**.46**	**.27**							
Physical Fitness (PF)	**.12**	.05	**.22**						
Teamwork (TEAM)	.03	**.19**	-.07	**-.39**					
Satisfaction with the Army (ASat)	**.19**	**.10**	**.34**	**.59**	**-.46**				
Perceived Fit with Army (AFit)	**.21**	**.26**	**.33**	**.62**	**-.42**	**.86**			
Attrition Cognitions (ACog)	**-.24**	**-.12**	**-.39**	**-.42**	**.55**	**-.64**	**-.61**		
Career Intentions (CInt)	**.14**	**.12**	**.38**	**.57**	**-.34**	**.68**	**.75**	**-.55**	
Future Army Affect (FAA)	**.25**	**.09**	**.35**	**.41**	**-.52**	**.57**	**.62**	**-.50**	**.48**

Note. n = 682. Statistically significant correlations are bolded ($p < .05$, two-tailed). FXP = Dyad composite for Future Expected Performance, GTP = Dyad composite for General Technical Proficiency, AE = Dyad composite for Achievement and Effort, PF = Dyad composite for Physical Fitness, TEAM = Dyad composite for Teamwork, ASat = Dyad composite for Satisfaction with the Army, AFit = Dyad composite for Perceived Army Fit, ACog = Dyad composite for Attrition Cognitions, CInt = Dyad composite for Career Intentions, and FAA = Dyad composite for Future Army Affect.

Tables 8.8 and 8.9 provide details about the structure of the EDCs. Table 8.8 shows which dyads make up the EDC for each of the 10 criterion composites. Table 8.9 presents the dyad as the focus, linking each dyad that appears in a WSI EDC with the criterion measure(s) that it predicts. The table shows that 34 of the potential 120 unique dyads appeared in the EDCs. Of these 34, 24 appeared in just one EDC, 5 appeared in two, 4 appeared in three, and 1 (signifying that the respondent ranked Leadership Orientation higher than Innovation) appeared in five EDCs. Such results provide partial evidence for the discriminant validity of the performance dimensions.

Table 8.8. Dyads that Contribute to Each Empirical Dyad Composite

Performance Criteria	WSI Dyad Components
Future Expected Performance	Attention to Detail ranked higher than Cooperation
	Dependability ranked higher than Independence
	Independence ranked higher than Social Orientation
	Leadership Orientation ranked higher than Energy
General Technical Proficiency	Attention to Detail ranked higher than Cooperation
	Independence ranked higher than Energy
	Leadership Orientation ranked higher than Energy
	Stress Tolerance ranked higher than Concern for Others
	Stress Tolerance ranked higher than Initiative
Achievement and Effort	Achievement and Effort ranked higher than Attention to Detail
	Achievement and Effort ranked higher than Self-Control
	Attention to Detail ranked higher than Cooperation
	Dependability ranked higher than Energy
	Leadership Orientation ranked higher than Innovation
Physical Fitness	Energy ranked higher than Innovation
	Initiative ranked higher than Self-Control
	Innovation ranked higher than Concern for Others
	Leadership Orientation ranked higher than Independence
Teamwork	Concern for Others ranked higher than Adaptability/Flexibility
	Independence ranked higher than Energy
Satisfaction with the Army	Attention to Detail ranked higher than Independence
	Dependability ranked higher than Innovation
	Energy ranked higher than Cultural Tolerance
	Initiative ranked higher than Self-Control
	Leadership Orientation ranked higher than Innovation
	Self-Control ranked higher than Independence
	Social Orientation ranked higher than Concern for Others
	Cultural Tolerance ranked higher than Concern for Others

Table 8.8. (Continued)

Attitudinal Criteria	WSI Dyad Components
Perceived Army Fit	Attention to Detail ranked higher than Concern for Others
	Energy ranked higher than Cultural Tolerance
	Initiative ranked higher than Self-Control
	Leadership Orientation ranked higher than Innovation
	Self-Control ranked higher than Independence
	Stress Tolerance ranked higher than Innovation
	Cultural Tolerance ranked higher than Concern for Others
Attrition Cognitions	Concern for Others ranked higher than Achievement and Effort
	Cooperation ranked higher than Initiative
	Independence ranked higher than Energy
	Innovation ranked higher than Dependability
	Cultural Tolerance ranked higher than Stress Tolerance
Career Intentions	Dependability ranked higher than Concern for Others
	Energy ranked higher than Self-Control
	Initiative ranked higher than Independence
	Leadership Orientation ranked higher than Innovation
	Stress Tolerance ranked higher than Innovation
Future Army Affect	Achievement and Effort ranked higher than Social Orientation
	Adaptability/Flexibility ranked higher than Social Orientation
	Energy ranked higher than Independence
	Initiative ranked higher than Social Orientation
	Leadership Orientation ranked higher than Innovation
	Social Orientation ranked higher than Concern for Others
	Cultural Tolerance ranked higher than Concern for Others

Table 8.9. Mapping of Dyads onto Criteria

Dyad	Criterion Variable
Achievement and Effort ranked higher than Attention to Detail	Achievement and Effort
Achievement and Effort ranked higher than Self-Control	Achievement and Effort
Achievement and Effort ranked higher than Social Orientation	Future Army Affect
Adaptability/Flexibility ranked higher than Social Orientation	Future Army Affect
Attention to Detail ranked higher than Concern for Others	Perceived Army Fit
Attention to Detail ranked higher than Cooperation	Achievement and Effort Future Expected Performance General Technical Proficiency
Attention to Detail ranked higher than Independence	Satisfaction with the Army
Concern for Others ranked higher than Achievement and Effort	Attrition Cognitions
Concern for Others ranked higher than Adaptability/Flexibility	Teamwork
Cooperation ranked higher than Initiative	Attrition Cognitions
Dependability ranked higher than Concern for Others	Career Intentions
Dependability ranked higher than Energy	Achievement and Effort
Dependability ranked higher than Independence	Future Expected Performance
Dependability ranked higher than Innovation	Satisfaction with the Army
Energy ranked higher than Independence	Future Army Affect
Energy ranked higher than Innovation	Physical Fitness
Energy ranked higher than Cultural Tolerance	Perceived Army Fit Satisfaction with the Army
Independence ranked higher than Energy	Attrition Cognitions General Technical Proficiency Teamwork
Independence ranked higher than Social Orientation	Future Expected Performance
Initiative ranked higher than Independence	Career Intentions
Initiative ranked higher than Self-Control	Physical Fitness Perceived Army Fit Satisfaction with the Army

Table 8.9. (Continued)

Dyad	Criterion Variable
Initiative ranked higher than Social Orientation	Future Army Affect
Innovation ranked higher than Concern for Others	Physical Fitness
Innovation ranked higher than Dependability	Attrition Cognitions
Leadership Orientation ranked higher than Energy	Future Expected Performance General Technical Proficiency
Leadership Orientation ranked higher than Independence	Physical Fitness
Leadership Orientation ranked higher than Innovation	Achievement and Effort Career Intentions Future Army Affect Perceived Army Fit Satisfaction with the Army
Self-Control ranked higher than Independence	Perceived Army Fit Satisfaction with the Army
Social Orientation ranked higher than Concern for Others	Future Army Affect Satisfaction with the Army
Stress Tolerance ranked higher than Concern for Others	General Technical Proficiency
Stress Tolerance ranked higher than Initiative	General Technical Proficiency
Stress Tolerance ranked higher than Innovation	Career Intentions Perceived Army Fit
Cultural Tolerance ranked higher than Concern for Others	Future Army Affect Perceived Army Fit Satisfaction with the Army
Cultural Tolerance ranked higher than Stress Tolerance	Attrition Cognitions

Validity Results

The previous section provided basic descriptive statistics for the WSI scores. In this section, we examine the degree to which the WSI full scores and EDCs correlate with the 10 Select21 criteria. Table 8.10 shows raw (i.e., uncorrected) criterion-related validity estimates for WSI scores in the total CV sample (i.e., Waves 1 and 2 combined). The table also has a row containing correlations between gender and the criteria. This row serves as a reference point, helping determine the degree to which the WSI scores might be serving as little more than a proxy for gender.

Examination of Table 8.10 suggests the following:

- Regarding the full scores, Concern for Others (score D) and Stress Tolerance (score O) were the best predictors. None of the full scores did a good job of predicting the set of performance criteria; the bulk of predictive validity was observed for the attitudinal criteria.
- The EDC composites correlated reasonably well (r =.14 to .27) with the target performance criteria (i.e., the criteria against which the EDCs were keyed) and quite well with the target attitudinal criteria (r = .29 to .39). The EDCs clearly serve as more than a gender proxy.
- Gender correlated moderately with half the criteria (correlations ranging from .08 to .11 in absolute value) and weakly with four others (General Technical Proficiency, Physical Fitness, Perceived Fit with Army, Career Intentions). The highest correlation was .16 with Achievement and Effort. These findings, combined with those for the EDC composites, clearly indicate that the EDCs serve as more than a proxy for gender.
- Although each EDC correlated more highly with its target criterion than any other EDC, it does not follow that a given EDC correlated more highly with its target criterion than with any other criterion. For example, no EDC predicted Teamwork better than the Teamwork EDC (r = .14); the other EDCs correlated only -.02 to .09 with Teamwork. Nevertheless, the Teamwork EDC correlated higher than .14 with all five attitudinal criteria.

Table 8.11 contains corrected validity estimates—that is, the raw validity estimates presented in Table 8.10 after correction for criterion unreliability and range restriction. Examination of Table 8.11 suggests the following:

- Correlations with the performance criteria were generally in the mid- to upper- .20s, although the magnitude of the correlation for General Technical Proficiency was more in line with those seen for the attitudinal criteria.
- Correlations with the attitudinal criteria ranged through the .30s to the lower .40s, with the best prediction obtained for (a) Perceived Fit with the Army and (b) Satisfaction with the Army.
- The attitudinal criteria were predicted well not only by their targeted EDCs, but also by the EDCs designed for other attitudinal criteria. This phenomenon did not apply uniformly to the performance criteria. For example, although the corrected validity estimate for General Technical Proficiency was .39 for the General Technical Proficiency EDC, only one other EDC achieved a validity estimate greater than .12 (the Expected Future Performance EDC, with a validity estimate of .23 for General Technical Proficiency). Note, however, that Achievement and Effort and Physical Fitness did show some "cross-EDC" prediction.
- The full scores continued to show respectable predictive validity, although their validity estimates were smaller than those for the EDCs.

Table 8.10. Uncorrected Criterion-Related Validity Estimates for WSI Scores in the Full CV Sample

		Performance Criteria					Attitudinal Criteria				
	WSI Score	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
Full Score	D: Concern for Others	**-.18**	-.06	**-.15**	.06	-.04	**-.21**	**-.23**	**.18**	**-.13**	**-.18**
	F: Dependability	.03	**.13**	.00	.04	.08	**.11**	.09	**-.11**	**.14**	.04
	H: Independence	.08	-.02	-.05	.02	.01	**-.15**	**-.13**	**.10**	-.08	-.07
	J: Innovation	-.03	**-.13**	-.08	-.02	-.04	**-.17**	**-.16**	.09	**-.17**	-.09
	K: Leadership Orientation	**.15**	.10	.08	.02	.09	.04	**.11**	-.04	.04	**.10**
	O: Stress Tolerance	**.11**	.00	.05	-.02	.02	**.13**	**.15**	**-.11**	**.15**	**.13**
	Gender (Female = 1)	*.02*	***.16***	*-.03*	*.08*	*.10*	*-.09*	*-.03*	*.09*	*.00*	***-.11***
Empirical Dyad Composite	General Technical Proficiency	**.27**	**.12**	.07	.08	**.15**	.05	**.10**	-.07	.03	.06
	Achievement and Effort	**.11**	**.27**	**.12**	.06	**.14**	**.18**	**.22**	**-.17**	**.12**	**.12**
	Physical Fitness	**.11**	**.13**	**.24**	-.02	.08	**.23**	**.28**	**-.19**	**.17**	**.16**
	Teamwork	.04	.01	**-.12**	**.14**	.02	**-.20**	**-.18**	**.17**	**-.15**	**-.17**
	Expected Future Performance	**.19**	**.19**	.09	.09	**.20**	.09	**.11**	**-.11**	.03	.07
	Satisfaction with the Army	.09	**.19**	**.19**	-.01	.06	**.38**	**.37**	**-.27**	**.25**	**.25**
	Perceived Fit with the Army	**.12**	**.17**	**.21**	-.02	.08	**.35**	**.39**	**-.24**	**.27**	**.28**
	Attrition Cognitions	-.06	-.10	**-.12**	.06	-.03	**-.24**	**-.28**	**.29**	**-.19**	**-.17**
	Career Intentions	.06	**.18**	**.16**	.00	.07	**.26**	**.32**	**-.17**	**.30**	**.23**
	Future Army Affect	.06	.09	**.16**	-.08	.04	**.26**	**.29**	**-.23**	**.21**	**.34**

Note. n = 498-645. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect. Validity estimates for Gender provided for reference purposes only. Statistically significant correlations are bolded ($p < .01$, one-tailed). Boxed/underscored correlations denote validity estimates for criteria to which the empirical dyad composites were keyed.

Table 8.11. Corrected Criterion-Related Validity Estimates for WSI Scores in the Full CV Sample

		Performance Criteria					Attitudinal Criteria				
	WSI Score	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
Full Score	D: Concern for Others	-.27	-.10	-.16	.08	-.11	-.21	-.25	.24	-.11	-.18
	F: Dependability	.03	.14	.00	.06	.11	.12	.10	-.13	.14	.04
	H: Independence	.17	.01	-.04	.05	.07	-.17	-.15	.09	-.10	-.09
	J: Innovation	.03	-.12	-.08	-.02	-.01	-.19	-.18	.09	-.18	-.11
	K: Leadership Orientation	.17	.11	.09	.04	.12	.04	.12	-.05	.04	.10
	O: Stress Tolerance	.16	.03	.05	-.03	.06	.14	.16	-.16	.14	.13
Empirical Dyad Composite	General Technical Proficiency	.39	.18	.08	.16	.25	.04	.11	-.12	.01	.05
	Achievement and Effort	.12	.29	.12	.09	.18	.19	.24	-.20	.13	.12
	Physical Fitness	.11	.13	.24	-.04	.10	.25	.31	-.22	.18	.18
	Teamwork	.05	.01	-.12	.24	.03	-.22	-.20	.20	-.16	-.18
	Expected Future Performance	.23	.22	.09	.15	.27	.09	.13	-.14	.02	.06
	Satisfaction with the Army	.06	.19	.19	-.03	.04	.40	.41	-.31	.27	.27
	Perceived Fit with Army	.10	.17	.21	-.04	.08	.38	.43	-.27	.29	.30
	Attrition Cognitions	-.09	-.12	-.12	.10	-.05	-.25	-.31	.35	-.20	-.17
	Career Intentions	.03	.18	.16	-.02	.06	.28	.36	-.19	.32	.25
	Future Army Affect	.06	.09	.17	-.14	.05	.28	.32	-.27	.22	.36

Note. n = 498-645. Validity estimates were first corrected for unreliability in the criterion and then for indirect range restriction resulting from selection on the AFQT. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect. Boxed/underscored correlations denote validity estimates for criteria to which the empirical dyad composites were keyed.

Cross-Validation of Composites

As part of a preliminary analysis, EDCs were created for the Wave 1 sample. Given the empirical keying of the WSI, a natural question regards how well such empirical composites cross-validate when applied to a different sample. The Wave 2 sample gave us the opportunity to answer this question.

Table 8.12 shows criterion-related validity estimates for WSI EDCs in the Wave 1 and Wave 2 samples. (The table also contains validity estimates for the total CV sample for comparison.) Unlike Table 8.10, this table contains validity estimates based on EDCs that were constructed based on the Wave 1 sample data *only*. Therefore, the criterion-related validity estimates for the EDCs in the Wave 2 sample represent cross-validities (i.e., criterion-related validity estimates obtained by applying Wave 1 EDCs to Wave 2 data; see Figure 8.1 for a comparison of the WSI analysis strategies).

Primary Analysis

- Derive Empirical Dyad Composites (EDCs) for each criterion variable
- Analysis sample = Total CV sample (i.e., Waves 1 and 2 combined)

Cross-Validation Analysis

- Calculate correlations in the Wave 2 sample between (a) criterion scores and (b) EDCs obtained from Wave 1 analysis
- Analysis sample = Wave 2 sample

Figure 8.1. Comparison of primary and cross-validation analysis strategies.

Table 8.12. Criterion-Related Validity Estimates for WSI Empirical Dyad Composites in the Wave 1, Wave 2, and Full CV Samples

	Performance Criteria					Attitudinal Criteria				
Sample/Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
Uncorrected Estimates										
Total Sample	.27	.27	.24	.14	.20	.38	.39	.29	.30	.34
Wave 1	.25	.28	.30	.08	.21	.37	.37	.27	.27	.36
XVal Sample	**.19**	**.18**	.01	**.19**	**.14**	**.27**	**.38**	**.23**	**.21**	**.20**
Corrected Estimates										
Total Sample	.39	.29	.24	.24	.27	.40	.43	.35	.32	.36
Wave 1	.37	.31	.31	.16	.28	.39	.42	.33	.28	.38
XVal Sample	.34	.22	.01	.30	.29	.28	.41	.27	.19	.20

Note. n_{Wave1} = 359 (AE criterion), n_{Wave1} = 496 (all other performance criteria), n_{Wave1} = 462-478 (attitudinal criteria). n_{Wave2} = 139 (AE criterion), n_{Wave2} = 149 (all other performance criteria), n_{Wave2} = 155-157 (attitudinal criteria). Uncorrected estimates for the XVal (i.e., cross-validation) are correlations of Wave 1 EDCs with Wave 2 criteria and hence constitute cross-validity estimates. Corrected validity estimates were first corrected for unreliability in the criterion and then for indirect range restriction resulting from selection on the AFQT. Statistically significant uncorrected cross-validation validity estimates are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 8.12 shows that, for the most part, the Wave 1 EDCs retained their validity in the Wave 2 data. The decrement in corrected cross-validity for the majority of criteria ranged from 3 to 11 correlation points. Note, however, that the pattern of results differs between the performance criteria and the attitudinal criteria. For the EDCs targeting Teamwork and Expected Future Performance, the corrected cross-validity estimates (r = .30 and .29, respectively) were actually higher than the Wave 1 estimates (r = .16 and .28, respectively). The remaining two criteria (Physical Fitness, Future Army Affect) provided more disappointing results, with the corrected cross-validity of the former falling to near zero and that of the latter dropping 18 correlation points from .38 to .20. Even so, the corrected cross-validity estimates ranged from about .20 to about .40—certainly of sufficient magnitude to provide reasonable utility should these values be obtained in an operational setting.

Incremental Validity Estimates

The previous section presented evidence for the criterion-related validity of the WSI. In this section, the focus shifts to the incremental validity of the WSI scores over the AFQT.

Table 8.13 shows incremental validity estimates for the select WSI full scores and the 10 EDCs in the total CV sample. The estimates in the table show that the WSI offered substantial incremental validity over the AFQT for predicting the attitudinal criteria. This finding is not surprising given the general lack of validity of the AFQT for predicting attitudinal criteria and the strength of the WSI for predicting attitudinal criteria (see Tables 8.10 and 8.11). With regard to the performance criteria, the incremental validity of the WSI composites over the AFQT was notable for all criteria but especially so for Achievement and Effort, Physical Fitness, and Teamwork. The results for these latter three criteria are in line with expectations, given that predictors that assess motivation-related determinants of performance (such as the WSI) should have the best chance for incremental validity for performance criteria having more of a "will-do" component. Yet, the WSI provided incremental validity even to the "can-do" criterion General Technical Proficiency (specifically, an increase of 5 correlation points)—this despite the strong relation between AFQT and General Technical Proficiency (corrected validity estimate of .55). Finally, the significant increment for Physical Fitness could be due in part to attitudinal variables that relate to both the WSI and Physical Fitness (e.g., Satisfaction with the Army). Further, we would expect measures of cognitive ability such as the AFQT to have little to do with physical fitness performance (and the corrected value is only .06), thus increasing the potential for incremental validity.

Table 8.13. Incremental Validity Estimates for WSI Scores in the Full Sample

		Uncorrected Incremental Validity Estimates									
		Performance Criteria					Attitudinal Criteria				
	Composite/Score	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	AFQT	**.32**	**.16**	.01	.05	**.19**	-.03	.00	**-.11**	-.07	-.05
Full Score	D: Concern for Others	**.03**	.00	**.14**	.03	.00	**.19**	**.23**	**.09**	**.08**	**.15**
	F: Dependability	.00	**.05**	.00	.01	**.02**	.08	.08	.05	**.08**	.01
	H: Independence	.00	.01	.04	.00	.00	**.12**	**.13**	**.05**	.03	.03
	J: Innovation	.01	**.06**	**.07**	.01	.01	**.14**	**.16**	.04	**.10**	.05
	K: Leadership Orientation	**.03**	**.03**	**.07**	.00	**.02**	.02	.10	.01	.01	.06
	O: Stress Tolerance	.01	.00	.03	.01	.00	**.11**	**.14**	.04	**.10**	**.09**

Table 8.13. (Continued)

		Uncorrected Incremental Validity Estimates									
		Performance Criteria					Attitudinal Criteria				
	Composite/Score	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
EDC Composite	General Technical Proficiency (GTP)	**.06**	.02	.06	.04	**.03**	.06	.11	.04	.01	.03
	Achievement and Effort (AE)	**.02**	**.16**	**.10**	.02	**.05**	.03	.10	.01	.01	.04
	Physical Fitness (PF)	**.02**	**.05**	**.22**	.00	**.02**	**.15**	**.21**	**.09**	**.07**	.07
	Teamwork (TEAM)	.00	.00	**.11**	**.10**	.00	**.20**	**.28**	**.11**	**.11**	**.12**
	Future Expected Performance (FXP)	**.04**	**.08**	**.07**	**.05**	**.08**	**.17**	**.17**	**.09**	**.09**	**.12**
	Satisfaction with the Army (ASat)	**.02**	**.10**	**.18**	.00	.01	**.34**	**.36**	**.19**	**.18**	**.20**
	Perceived Army Fit (AFit)	**.03**	**.08**	**.20**	.00	**.02**	**.32**	**.38**	**.16**	**.21**	**.23**
	Attrition Cognitions (ACog)	.00	**.03**	**.10**	.03	.00	**.21**	**.27**	**.19**	**.14**	**.13**
	Career Intentions (CInt)	**.01**	**.09**	**.14**	.00	**.02**	**.23**	**.32**	**.10**	**.23**	**.18**
	Future Army Affect (FAA)	.01	.02	**.15**	**.04**	.00	**.23**	**.28**	**.14**	**.15**	**.29**

		Corrected Incremental Validity Estimates									
		Performance Criteria					Attitudinal Criteria				
	Composite/Score	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	AFQT	.55	.27	.02	.14	.38	-.05	-.01	-.21	-.12	-.09
Full Score	D: Concern for Others	.02	.00	.14	.04	.00	.18	.25	.08	.07	.13
	F: Dependability	.00	.04	.00	.01	.02	.07	.09	.04	.07	.01
	H: Independence	.00	.01	.03	.00	.00	.12	.14	.04	.02	.02
	J: Innovation	.00	.05	.07	.01	.01	.13	.18	.04	.08	.04
	K: Leadership Orientation	.02	.02	.07	.00	.02	.02	.11	.00	.01	.05
	O: Stress Tolerance	.01	.00	.03	.01	.00	.10	.16	.03	.08	.08
EDC Composite	General Technical Proficiency (GTP)	.05	.02	.06	.05	.03	.03	.11	.01	.01	.03
	Achievement and Effort (AE)	.01	.13	.10	.03	.04	.15	.23	.08	.06	.06
	Physical Fitness (PF)	.01	.04	.23	.00	.02	.20	.31	.10	.09	.10
	Teamwork (TEAM)	.00	.00	.10	.14	.00	.17	.19	.08	.08	.11
	Future Expected Performance (FXP)	.03	.06	.07	.06	.07	.06	.12	.03	.00	.03
	Satisfaction with the Army (ASat)	.01	.08	.18	.00	.01	.35	.40	.18	.16	.19
	Perceived Army Fit (AFit)	.02	.06	.20	.00	.02	.32	.43	.15	.18	.22
	Attrition Cognitions (ACog)	.00	.02	.10	.04	.00	.21	.30	.19	.12	.11
	Career Intentions (CInt)	.01	.07	.14	.00	.01	.22	.35	.09	.21	.17
	Future Army Affect (FAA)	.00	.02	.15	.06	.00	.23	.32	.13	.13	.28

Note. n = 595-611. Cell values for the AFQT represent zero-order correlations between AFQT and the given criterion and are shown for reference. Uncorrected estimates reflect the difference between the multiple R obtained when regressing the criterion on both the given predictor and AFQT, and the R obtained when regressing the criterion on AFQT only. Statistically significant incremental validities are bolded ($p < .05$, one-tailed). Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Boxed/underscored correlations denote incremental validity estimates for criteria to which the empirical dyad composites were keyed.

Subgroup Differences

Earlier in this chapter, we presented by gender both the mean ranks (Table 8.4) and the rank orders of those mean ranks (Table 8.5) for the WSI full scores. In this section, we present subgroup data for the WSI EDCs. Tables 8.14 and 8.15 show means for the EDCs by gender and race/ethnicity, respectively. For the gender comparisons, Table 8.14 shows that 3 of the 10 WSI EDCs showed significant effect sizes, indicating that women scored significantly higher on the EDCs for Teamwork and Attrition Cognitions but significantly lower on the EDC for predicting Perceived Fit with the Army. The overall magnitude of these effect sizes was moderate, and two of the three EDCs mirror significant mean differences on the criteria themselves (the exception is Perceived Army Fit).

Table 8.14. Final WSI Empirical Dyad Composite Scores by Gender

		Female		Male	
WSI EDC	d_{FM}	*M*	*SD*	*M*	*SD*
General Technical Proficiency (GTP)	-0.23	2.33	1.07	2.61	1.22
Achievement and Effort (AE)	0.20	3.09	1.28	2.86	1.15
Physical Fitness (PF)	-0.24	1.89	0.98	2.11	0.92
Teamwork (TEAM)	**0.27**	1.08	0.70	0.89	0.70
Future Expected Performance (FXP)	-0.10	2.20	0.84	2.29	0.89
Satisfaction with the Army (ASat)	-0.22	3.75	1.98	4.10	1.61
Perceived Army Fit (AFit)	**-0.33**	3.05	1.70	3.52	1.42
Attrition Cognitions (ACog)	**0.38**	2.95	1.25	2.48	1.23
Career Intentions (CInt)	0.07	2.36	1.33	2.27	1.22
Future Army Affect (FAA)	-0.25	3.44	1.75	3.80	1.45

Note. n_{Male} = 580. n_{Female} = 64. $d_{F\text{-}M}$ = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 8.15. Final WSI Empirical Dyad Composite Scores by Race/Ethnic Group

			Black		White		Hispanic		White Non-Hispanic	
WSI EDC	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
GTP	**-0.48**	**-0.23**	2.10	1.16	2.68	1.21	2.45	1.14	2.73	1.22
AE	-0.08	0.05	2.81	1.10	2.91	1.19	2.96	1.19	2.90	1.18
PF	**-0.30**	0.07	1.86	0.86	2.14	0.94	2.19	0.90	2.12	0.94
TEAM	0.15	0.02	0.99	0.72	0.89	0.69	0.89	0.67	0.88	0.71
FXP	-0.13	0.03	2.19	0.88	2.31	0.90	2.32	0.88	2.30	0.90
ASat	-0.20	-0.06	3.77	1.57	4.11	1.66	4.06	1.64	4.15	1.68
AFit	**-0.29**	-0.05	3.10	1.31	3.53	1.48	3.48	1.40	3.56	1.51
ACog	**0.28**	0.19	2.80	1.12	2.45	1.24	2.63	1.12	2.39	1.27
CInt	**-0.21**	0.01	2.06	1.15	2.33	1.25	2.33	1.25	2.32	1.24
FAA	-0.20	0.02	3.53	1.32	3.83	1.51	3.85	1.35	3.81	1.57

Note. n_{Black} = 110. n_{White} = 469. $n_{Hispanic}$ = 127. $n_{White\ Non\text{-}Hispanic}$ = 366. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites)/*SD* of Whites. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, ACog = Attrition Cognitions, CInt = Career Intentions, FAA = Future Army Affect.

Table 8.15 shows that whereas only one of the WSI EDCs exhibited a significant effect size for the Hispanic/White Non-Hispanic comparison, 5 of the 10 EDCs (those for General Technical Proficiency, Physical Fitness, Perceived Fit with the Army, Attrition Cognitions, and Career Intentions) had significantly different effect sizes for the Black/White comparison. Relative to Whites, Black Soldiers scored lower on four of these five composites (the exception being Attrition Cognitions). Again, all of these effect sizes were in the moderate range except for the General Technical Proficiency effect size of -0.48, which would be considered large. Note, however, that this effect size might simply be mirroring the mean difference observed on General Technical Proficiency for this sample: the criterion effect size itself is -0.45. Nevertheless, these effect sizes are somewhat unexpected and merit further investigation.

Differential Prediction

Tables 8.16 – 8.18 present the results of differential prediction analyses for the WSI empirical dyad composites. We performed three subgroup comparisons—one involving gender (female/male) and two involving race/ethnicity (Black/White and Hispanic/White non-Hispanic, respectively). Overall, the results indicated minor intercept bias (primarily for gender) and very little differential prediction (i.e., slope bias—only 2 of 30 tests were significant). With regard to intercept bias, a common regression line would lead to *underprediction* for the focal group in five of the seven cases; *overprediction* would occur only for females on Future Army Affect and for Blacks on General Technical Proficiency. With regard to slope bias, the two instances (for females on General Technical Proficiency, for Hispanics on Future Army Affect) both showed lesser predictive validity for the EDC in the focal group than in the referent group. Specifically, the General Technical Proficiency EDC/criterion correlation was 37 points lower for females than for males (-.06 and .30, respectively—the discrepancy is due to rounding); similarly, the Future Army Affect EDC/criterion correlation was 29 points lower for Hispanics than for White Non-Hispanics (.12 and .41, respectively).

Table 8.16. Differential Prediction Results for Final WSI Empirical Dyad Composites by Gender

	WSI Empirical Dyad Composite				
		WSI *b*		*r* by Gender	
Criterion	Gender *b*	M	F	M	F
Performance Criteria					
General Technical Proficiency	0.00	**0.16**	**-0.04**	**0.30**	-0.06
Achievement and Effort	**0.23**	0.14	0.09	**0.27**	0.20
Physical Fitness	-0.04	0.18	0.10	**0.25**	0.15
Teamwork	0.14	0.08	0.10	**0.14**	0.16
Future Expected Performance	**0.22**	0.15	-0.01	**0.22**	-0.02
Attitudinal Criteria					
Satisfaction with the Army	-0.18	0.29	0.19	**0.38**	**0.31**
Perceived Army Fit	0.04	0.31	0.35	**0.37**	**0.49**
Attrition Cognitions	**0.27**	0.29	0.09	**0.30**	0.09
Career Intentions	-0.05	0.32	0.35	**0.29**	**0.32**
Future Army Affect	**-0.28**	0.30	0.41	**0.31**	**0.53**

Note. $n_{\text{Regression}}$ = 497-644. n_{Male} = 437-580. n_{Female} = 60-64. Gender *b* = Unstandardized regression weight for gender (0 = male, 1 = female). WSI *b* = Unstandardized regression weight for the given WSI composite for males and females. *r* by Gender = Correlation between the given WSI composite and the given criterion for each gender. Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Regression weights for males and females are bolded if the WSI-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 8.17. Differential Prediction Results for Final WSI Empirical Dyad Composites by Race

	WSI Empirical Dyad Composite				
		WSI *b*		*r* by Race	
Criterion	Race *b*	W	B	W	B
Performance Criteria					
General Technical Proficiency	**-0.20**	0.16	0.05	**0.30**	0.11
Achievement and Effort	-0.09	0.12	0.23	**0.23**	**0.38**
Physical Fitness	0.07	0.19	0.13	**0.26**	0.17
Teamwork	-0.01	0.08	0.14	**0.14**	**0.24**
Future Expected Performance	-0.13	0.13	0.10	**0.19**	0.17
Attitudinal Criteria					
Satisfaction with the Army	0.03	0.30	0.18	**0.40**	**0.23**
Perceived Army Fit	-0.02	0.33	0.16	**0.42**	**0.20**
Attrition Cognitions	**0.28**	0.28	0.14	**0.29**	0.15
Career Intentions	0.09	0.33	0.32	**0.30**	**0.29**
Future Army Affect	-0.14	0.33	0.31	**0.35**	**0.30**

Note. $n_{\text{Regression}}$ = 452-579. n_{White} = 363-469. n_{Black} = 89-110. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black). WSI *b* = Unstandardized regression weight for the given WSI composite for Whites and Blacks. *r* by Race = Correlation between the given WSI composite and the given criterion for each race. Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Regression weights for Whites and Blacks are bolded if the WSI-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 8.18. Differential Prediction Results for Final WSI Empirical Dyad Composites by Ethnic Group

	WSI Empirical Dyad Composite				
		WSI *b*		*r* by Eth	
Criterion	Eth *b*	W	H	W	H
Performance Criteria					
General Technical Proficiency	-0.05	0.17	0.10	**0.32**	**0.19**
Achievement and Effort	0.04	0.10	0.15	**0.19**	**0.32**
Physical Fitness	0.05	0.18	0.25	**0.24**	**0.33**
Teamwork	0.11	0.11	0.01	**0.18**	0.02
Future Expected Performance	0.03	0.12	0.16	**0.17**	**0.26**
Attitudinal Criteria					
Satisfaction with the Army	0.13	0.32	0.22	**0.44**	**0.30**
Perceived Army Fit	0.14	0.36	0.23	**0.46**	**0.28**
Attrition Cognitions	-0.02	0.27	0.30	**0.29**	**0.29**
Career Intentions	-0.05	0.35	0.37	**0.31**	**0.39**
Future Army Affect	**0.29**	**0.36**	**0.13**	**0.41**	0.12

Note. $n_{\text{Regression}}$ = 378-493. $n_{\text{White non-Hispanic}}$ = 286-366. $n_{Hispanic}$ = 92-127. Eth *b* = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). WSI *b* = Unstandardized regression weight for the given WSI composite for White non-Hispanics and Hispanics. *r* by Eth = Correlation between the given WSI composite and the given criterion for each race. Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Regression weights for White non-Hispanics and Hispanics are bolded if the WSI-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Discussion

The results of the analyses presented in this chapter suggest that the WSI could provide the Army with a reasonable option for assessing personality. Examination of the criterion-related validity of the WSI suggests it has substantial promise for predicting attitudinal criteria, which many view as precursors of attrition and re-enlistment behavior (Strickland, 2005). Results also indicate that the WSI has promise for predicting performance criteria above and beyond the AFQT—particularly for Achievement and Effort, Physical Fitness, and Teamwork performance. The findings with regard to the criterion-related validity of the WSI observed in this chapter compare favorably to those found in past Army research (Oppler, McCloy, & Campbell, 2001; Oppler, McCloy, Peterson, Russell, & Campbell, 2001; White, Young, & Rumsey, 2001). More importantly, the various scoring approaches to the EDCs enhance the WSI's resistance to response distortion, which gives the measure greater promise for retaining similar predictive power once employed in an operational setting. In addition, it seems to have a small "assessment footprint," with a mean completion time of just 4 minutes and 45 seconds; indeed, 90% of all Soldiers in the CV sample completed the WSI in fewer than 7 minutes.

Some issues and questions regarding the WSI remain. First, we have yet to obtain test-retest data through which we could estimate the reliability of the instrument. Second, the promising concurrent validation results cannot be fully embraced until the measure has been tested in an operational environment. Third, there were several instances of intercept bias with the WSI, which would lead to over- or under-prediction of the performance of targeted subgroups if a total-sample regression line were used. Note, however, that these biases are less a function of WSI than a reflection that subgroup differences exist on the criteria of interest (see Chapters 3 through 5). On the positive side, there were but two examples of differential prediction, which easily could have resulted from chance given that we examined 10 EDCs, in three separate subgroup analyses, leading to 30 such tests. Fourth, there are many other scoring approaches that could be attempted with the WSI. We were unable to explore as many of these as we would have liked. Hence, future research with the WSI should explore the predictive power of alternative scoring procedures.

Regarding future use of the WSI, our recommendations revolve around the questions and issues just raised. First, we advise administering the WSI in a test-retest reliability study. The re-test interval should be long enough to reduce memory effects and yet minimize the opportunity for changes in standing on the WSI personality traits. We believe that an interval of at least one month would be required, with a 2-month interval preferred. One caution, however, regards the population of interest—Army applicants. If the WSI were administered to applicants and then again 2 months later, the applicants should be those in the Delayed Entry Program. The experience of basic training and its concomitant inculcation of Army values could result in artificially low test-retest reliability estimates for respondents who take the test during Initial Entry Training (IET).

Second, we suggest that the WSI be administered experimentally in an operational selection setting. Although this chapter has clearly demonstrated the validity of the WSI for predicting criteria in a concurrent sample, the operational context is the touchstone and ultimate test of the measure's utility. Previous Army research has demonstrated that the magnitude of differences between the psychometric properties of non-cognitive measures administered in operational and concurrent contexts can be substantial (Knapp, Waters et al., 2002). Although we

have designed the WSI to withstand the demands of operational assessment, unexpected factors could act to erode the promising initial results we have reported to date.

Although the project developed several measures of person-environment fit (which will be described in subsequent chapters of this report), the WSI was developed as a POP-Hybrid measure—that is, a hybrid of both a personality assessment and an assessment of person-organization fit. Future research should explore this other feature of the WSI to determine its potential utility.

In sum, the WSI provides an intriguing mix of low demands for testing time, strong validity in a research-only sample, and several defenses against response distortion. Should the WSI withstand its remaining tests (reliability, validity in an operational setting), we believe it might well serve as a promising new measure for predicting important Army criteria.

CHAPTER 9: VALIDATION OF THE RATIONAL BIODATA INVENTORY (RBI)

Robert N. Kilcullen
U.S. Army Research Institute

Dan J. Putka and Rodney A. McCloy
HumRRO

Overview

Biodata tests measure the test-taker's prior behavior, experiences, and reactions to life events using multiple-choice questions. Meta-analyses of the selection literature show that biodata effectively predict a wide variety of performance criteria (e.g., ratings of overall performance, advancement potential, commendations, sales volume, bonuses), with typical estimated validities in the .30 to .40 range (Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). In addition to being useful as an initial selection screen, biodata instruments achieve similar validity estimates for predicting various supervisory and managerial performance criteria (Owens, 1976; Reilly & Chao, 1982).

Empirical scoring keys have often been used with biodata instruments. Unfortunately, these scoring strategies have serious drawbacks. They often show high validity initially but suffer substantial shrinkage across samples and over time (Schwab & Oliver, 1974; Walker, 1985; White & Kilcullen, 1992). In addition, item selection and scoring rubrics are often atheoretical, which makes it difficult to understand what constructs the test is measuring or why particular criterion groups respond differently to certain items (Mumford & Stokes, 1991).

Awareness of these problems has led to increasing interest in rational keying strategies. These typically involve identifying constructs likely to predict the criterion of interest and subsequently writing biographical items to measure those predictor constructs (e.g., Emotional Stability, Conscientiousness). Item response weights are rationally assigned based upon the expected relations between the responses and the underlying construct. The scored item responses are then summed to form scale scores that have substantive meaning. These scales typically show good convergent and discriminant validity with personality "marker" scales measuring the same attributes, and generally show less susceptibility to socially desirable responding compared to their personality-based counterparts (Kilcullen, White, Mumford, & Mack, 1995).

The potential advantages of rational keying include a greater theoretical understanding of the phenomenon under study (Mumford & Stokes, 1991; Mumford, Uhlman, & Kilcullen, 1992). Additionally, rational keys typically yield criterion-related validity estimates that are comparable to those achieved with cross-validated empirical keys (Schoenfeldt, 1989; Uhlman, Reiter-Palmon, & Connelly, 1990) and tend to produce more stable validity estimates over time (Clifton, Kilcullen, Reiter-Palmon, & Mumford, 1992; White & Kilcullen, 1992). For these reasons, the rational keying approach was chosen as the method for developing and scoring the Select21 biodata test, known as the Rational Biodata Inventory (RBI).

Instrument Description

Temperament constructs were targeted for measurement with the RBI based on a job analysis that targeted future-oriented Soldier competencies (Sager, Russell, Campbell, & Ford, 2005) as well as a review of the constructs measured by other biodata tests developed by ARI, particularly the Assessment of Right Conduct (ARC) and the Test of Adaptable Personality (TAP) – that have proven track records for predicting both counterproductive behavior and job performance in the Army (Kilcullen, Goodwin, Chen, Wisecarver, & Sanders, 2002; Kilcullen, Mael, Goodwin, & Zazanis, 1999; Kilcullen, White, Sanders, & Hazlett, 2003).

Also included in the RBI is the Lie scale used in the TAP and the ARC to detect deliberate response distortion. Item scoring for the Lie scale is based on the endorsement of unlikely virtues. Previous research indicates that this scale shows good convergent and discriminant validity with a previously validated temperament scale measuring the same type of response distortion (Kilcullen et al., 1995). In addition, the RBI Lie scale demonstrates sensitivity to deliberate response distortion when respondents are instructed to fake good on the test (Kilcullen et al., 2005). Because the goal of Select21 is to develop selection tests for operational use where faking on self-report measures is a concern, the Lie scale in this research was used as one criterion for eliminating pilot items.

A detailed description of the development of the RBI is provided by Kilcullen, Putka, McCloy, and Van Iddekinge (2005). The version used in the concurrent validation had 101 items tapping 14 constructs, plus the Lie scale (see Figure 9.1)[25].

Method

Sample

The concurrent validation data collection yielded a sample of 719 Soldiers after eliminating cases with too much missing data or with indications of random responding to the predictor tests. An additional 31 cases (4% of the sample) were discarded because of elevated scores on the RBI Lie scale (which indicates socially desirable responding), leaving an RBI analysis sample of 688 Soldiers.

Analysis Approach

Elimination of 'high lie' responders in a concurrent validation study can make it easier to discern how predictor constructs relate to each other and to the criteria. However, in operational practice it would be difficult to justify eliminating applicants based on elevated Lie scale scores. Therefore, as a check, the analyses presented herein were also performed with the 31 "high lie" cases included in the sample. The results with the "high lie" cases included were virtually identical to the results presented herein.

[25] The Gratitude scale was originally targeted for deletion from the RBI before the concurrent validation study, due to its relatively low internal consistency. However, we decided to retain the Gratitude scale to allow for further assessment of its internal consistency and for assessment of its relationship to criteria (e.g., particularly criteria pertaining to teamwork and willingness to get along with others) that we expected it to predict.

Peer Leadership: Seeks positions of authority and influence. Comfortable with being in charge of a group. Willing to make tough decisions and accept responsibility for the group's performance. (6 items)
Cognitive Flexibility: Willingness to entertain new approaches to solving problems. Enjoys creating new plans and ideas. Initiates and accepts change and innovation. (8 items)
Achievement Orientation: The willingness to give one's best effort and to work hard towards achieving difficult objectives. (9 items)
Fitness Motivation: Degree of enjoyment from participating in physical exercise. Willingness to put in the time and effort to maintain good physical conditioning. (7 items, with 2 dropped from scoring in the concurrent validation)
Interpersonal Skills – Diplomacy: Being extroverted and outgoing. Able to make friends easily and establish rapport with strangers. Good at meeting/greeting people. (5 items)
Stress Tolerance: Ability to maintain one's composure under pressure. Remaining calm and in control of one's emotions instead of feeling anxious and worried. (11 items)
Hostility to Authority: Being suspicious of the motives and actions of legitimate authority figures. Viewing rules, regulations, and directives from higher authority as punitive and illegitimate. (7 items)
Self-Esteem: Feeling that one has successfully overcome work obstacles in the past and that one will continue to do so in the future. (6 items)
Cultural Tolerance: Willingness to work with people of different cultures. Being able to establish supportive work relationships with people with a variety of racial and ethnic backgrounds. (5 items)
Internal Locus of Control: The belief that one can exert influence over important events in order to control one's destiny. (8 items)
Army Identification: The degree of personal identification with, and intrinsic interest in becoming, a U.S. Army Soldier. (7 items)
Respect for Authority: Perceiving authority figures as having a positive influence on one's knowledge and skill development. (4 items)
Narcissism: Being excessively preoccupied with satisfying one's own needs and desires. (6 items)
Gratitude: Being appreciative of the help that one has received from others. (3 items)
Lie Scale: This scale is not a predictor scale. Its purpose is to detect and adjust for socially desirable responding. (7 items)

Figure 9.1. Rational Biodata Inventory (RBI) scales.

In this research, the Lie scale was used to adjust the RBI predictor scores such that the correlations of the predictor scores with the Lie score was no greater than r = .05. Once again, a separate set of analyses was conducted with the raw or unadjusted predictor scales, with the associated results nearly identical to those presented herein. This is not particularly surprising, because there is little motivation to fake in a concurrent validation setting. However, under operational conditions, the adjusted predictor scales may demonstrate higher validities to the extent that they more closely preserve the relative order of scores that are distorted when the scores of "fakers" and "non-fakers" are mixed.

Another issue to note is the artifactual criterion-related contamination for two RBI predictor scales – Fitness Motivation and Army Identification. The RBI Fitness Motivation scale measures intrinsic interest in maintaining a high degree of physical fitness. However, the scale also includes a few items relating to current level of fitness. These "level of fitness" items are contaminated with the Physical Fitness criterion used herein, which consists of ratings of the subject's physical fitness as well as scores on the Army Physical Fitness Test. The remedy adopted was to eliminate the level of fitness items from the RBI Fitness Motivation scale when the scale was used to predict the Physical Fitness criterion.

The second scale, RBI Army Identification, measures the degree of intrinsic interest in being a Soldier, or more generally emotional attachment to the Army. The attitudinal criteria used in this research incorporate a measure of Affective Commitment, which measures largely

the same construct. Thus, the estimated validities of the Army Identification scale for predicting the attitudinal criteria that are highly related to Affective Commitment (e.g., Satisfaction with the Army in General, Perceived Fit with the Army) were artificially inflated, and should be interpreted as such. These correlations are reported herein as an indicator of the construct validity of the Army Identification scale.

It is important to note that the contamination of the Fitness Motivation and Army Identification scales described above was the result of the concurrent validation design used herein, and not the result of an inherent defect in these scales. If these scales were administered to individuals prior to their entry into the Army, they could be evaluated free of concern from the contamination described above, and may well prove to be valid predictors of the criteria. Consistent with this notion, analyses of RBI Army Identification and RBI Fitness Motivation gathered from new recruits in the reception battalion in an earlier part of the Select21 effort significantly predicted early Soldier attrition (Putka & Bradley, 2006).

It is also important to note that the contamination issue described above is far less problematic when interpreting relations between the Army Identification scale and the job performance criteria, and between the Fitness Motivation scale and all of the attitudinal and non-Physical Fitness criteria.

Results

Psychometric Properties

Descriptive statistics and intercorrelations of the RBI scales are presented in Table 9.1. Acceptable internal consistency estimates for rational biodata scales are generally in the .60 and above range, given the heterogeneous nature of biodata items. A median RBI scale alpha of .73 was obtained, and all but the experimental Narcissism and Gratitude scales achieved the desired internal consistency estimate of at least .60.

Examination of the scale intercorrelations in Table 9.1 reveals six observed correlations at the .50 level or above, indicating strong overlap among the RBI scales of Peer Leadership, Cognitive Flexibility, and Achievement Orientation, as well as among these scales and the RBI Self Esteem scale. Other than these relations, the RBI scale intercorrelations were reasonably low, with only four observed correlations greater than .40. The moderate negative correlation between Hostility to Authority and Respect for Authority (r = -.20) suggests that these scales are not opposite ends of the same continuum.

Criterion-Related Validity Estimates

Table 9.2 reveals that all of the RBI predictor scales significantly predicted one or more of the performance criteria. Moreover, the directionality of the correlations made conceptual sense. Specifically, the RBI scales measuring desirable characteristics (e.g., Achievement Orientation) demonstrated positive correlations with the criteria, except for the negative criterion of Attrition Cognitions. On the other hand, the RBI Hostility to Authority scale (which measures an undesirable characteristic) was negatively correlated with the criteria except for the negative criterion of Attrition Cognitions. The only RBI scale not to show a consistent, expected pattern of correlations was the Narcissism scale.

Table 9.1. Descriptive Statistics, Reliability Estimates, and Intercorrelations among RBI Scales

Scale	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Peer Leadership	3.38	0.65	.72													
2. Cognitive Flexibility	3.42	0.71	**.53**	.82												
3. Achievement Orientation	3.30	0.62	**.56**	**.55**	.74											
4. Fitness Motivation	3.47	0.68	**.32**	**.26**	**.39**	.65										
5. Interpersonal Skills–Diplomacy	3.41	0.80	**.43**	**.27**	**.33**	**.29**	.76									
6. Stress Tolerance	2.86	0.51	-.02	.02	-.05	**.17**	**.27**	.68								
7. Hostility to Authority	2.76	0.67	.03	**-.16**	**-.23**	.03	-.03	**-.28**	.69							
8. Self-Esteem	3.88	0.59	**.53**	**.44**	**.50**	**.39**	**.41**	**.20**	-.06	.78						
9. Cultural Tolerance	3.74	0.78	**.29**	**.41**	**.30**	**.17**	**.42**	**.22**	**-.18**	**.34**	.76					
10. Internal Locus of Control	3.36	0.59	**.18**	**.27**	**.30**	**.23**	**.35**	**.36**	**-.30**	**.43**	**.29**	.69				
11. Army Identification	3.07	0.83	**.26**	**.18**	**.36**	**.27**	**.14**	.07	**-.21**	**.23**	**.10**	**.25**	.77			
12. Respect for Authority	3.33	0.67	**.25**	**.31**	**.53**	**.19**	**.19**	**-.08**	**-.20**	**.20**	**.18**	**.27**	**.37**	.65		
13. Narcissism	3.62	0.58	**.37**	**.20**	**.38**	**.20**	**.20**	**-.28**	**.22**	**.36**	.04	**.11**	**.10**	**.14**	.59	
14. Gratitude	3.82	0.73	**.27**	**.29**	**.38**	**.22**	**.38**	**.10**	**-.14**	**.35**	**.34**	**.34**	**.28**	**.38**	**.10**	.54
15. Lie Scale	0.05	0.09	.01	-.02	-.01	.06	.04	**.09**	-.03	.06	-.01	**.10**	.02	-.04	.01	.03

Note. n = 660-688. Internal consistency estimates are in the diagonal. Statistically significant correlations are bolded, $p < .05$ (two-tailed).

Table 9.2. Criterion-Related Validity Estimates for RBI Scales

	Performance Criteria					Attitudinal Criteria				
Scale	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
Peer Leadership	**.22**	**.19**	**.14**	.04	**.16**	.06	**.23**	**-.10**	**.10**	**.17**
Cognitive Flexibility	**.18**	**.16**	.04	**.09**	**.14**	**.09**	**.18**	**-.12**	**.08**	**.17**
Achievement Orientation	**.15**	**.26**	**.15**	**.09**	**.14**	**.28**	**.40**	**-.17**	**.19**	**.26**
Fitness Motivation	**.18**	**.15**	**.33**	.02	**.17**	**.22**	**.28**	**-.23**	**.13**	**.16**
Interpersonal Skills–Diplomacy	**.15**	**.13**	**.11**	.05	**.11**	**.19**	**.22**	**-.14**	**.08**	**.09**
Stress Tolerance	**.17**	**.11**	**.12**	.06	**.12**	**.24**	**.17**	**-.19**	**.08**	**.07**
Hostility to Authority	**-.19**	**-.30**	.02	**-.18**	**-.18**	**-.30**	**-.27**	**.20**	**-.13**	**-.10**
Self-Esteem	**.21**	**.23**	**.15**	**.08**	**.19**	**.14**	**.24**	**-.17**	**.13**	**.14**
Cultural Tolerance	**.08**	**.21**	.01	**.14**	**.12**	**.15**	**.21**	**-.15**	**.07**	**.17**
Internal Locus of Control	**.16**	**.24**	**.10**	**.08**	**.08**	**.32**	**.33**	**-.25**	**.14**	**.16**
Army Identification	**.19**	**.27**	**.15**	.04	**.14**	**.56**	**.69**	**-.47**	**.46**	**.48**
Respect for Authority	.03	**.15**	.03	.03	.04	**.31**	**.35**	**-.20**	**.22**	**.21**
Narcissism	-.04	-.04	**.12**	**-.08**	-.05	-.02	**.08**	.00	.03	**.12**
Gratitude	**.11**	**.19**	.02	**.07**	**.08**	**.27**	**.31**	**-.27**	**.12**	**.13**
Lie Scale	-.03	.06	.05	.02	.01	**.10**	**.12**	-.06	.04	.04
	Corrected Validity Estimates									
Peer Leadership	.32	.25	.15	.11	.27	.06	.26	-.16	.09	.17
Cognitive Flexibility	.34	.25	.05	.22	.28	.08	.19	-.20	.05	.16
Achievement Orientation	.20	.31	.16	.17	.21	.30	.44	-.22	.19	.27
Fitness Motivation	.22	.17	.34	.04	.23	.23	.31	-.29	.13	.17
Interpersonal Skills – Diplomacy	.19	.15	.12	.09	.16	.20	.25	-.17	.08	.10
Stress Tolerance	.25	.16	.13	.13	.21	.25	.19	-.26	.07	.07
Hostility to Authority	-.29	-.36	.01	-.33	-.29	-.31	-.30	.27	-.12	-.09
Self-Esteem	.30	.29	.16	.16	.29	.14	.27	-.23	.12	.14
Cultural Tolerance	.11	.24	.02	.24	.17	.16	.23	-.19	.07	.18
Internal Locus of Control	.24	.29	.11	.16	.16	.33	.36	-.33	.13	.16
Army Identification	.22	.30	.16	.08	.19	.59	.77	-.56	.47	.51
Respect for Authority	.04	.16	.04	.06	.06	.33	.39	-.24	.23	.23
Narcissism	-.06	-.05	.13	-.14	-.08	-.02	.08	.01	.03	.13
Gratitude	.18	.24	.02	.14	.15	.28	.34	-.35	.11	.13
Lie Scale	-.10	.03	.04	.01	-.03	.11	.13	-.04	.05	.05

Note. n = 487-508 (AE criterion), n = 634-658 (all other Performance criteria), n = 614-653 (Attitudinal criteria). Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Among the performance criteria, Expected Future Performance was predicted by 12 of the 14 RBI predictor scales, with Self Esteem (r = .19), Hostility to Authority (r = -.18) and Peer Leadership (r = .16) showing the highest validities (all $p < .05$). With respect to General Technical Performance, the best RBI predictors were Peer Leadership (r = .22), Self-Esteem (r = 21) and Army Identification (r = .19, all $p < .05$).

The estimated RBI scale validities obtained with the Achievement and Effort criterion were consistently higher compared to the other performance criteria. This finding is not surprising because the RBI scales reflect motivational constructs. Thirteen of the 14 RBI predictor scales were correlated with this outcome measure, with eight of the validity coefficients equal to or exceeding .19. Achievement and Effort was best predicted by Hostility to Authority (r = -.30), Army Identification (r = .27), Achievement Orientation (r = .26), Internal Locus of Control (r = .24), and Self Esteem (r = .23).

Nine of the 14 RBI predictor scales significantly predicted the Physical Fitness criterion. Not surprisingly, the best predictor was the adjusted RBI Fitness Motivation scale (r = .33, p < .05) with the "level of fitness" items removed. Other significant predictors of Physical Fitness included RBI Achievement, RBI Army Identification, and RBI Self-Esteem (all r = .15, p < .05).

Eight RBI scales predicted the Teamwork criterion, and the magnitude of these correlations was generally lower than with the other criteria, possibly because Teamwork had a relatively low internal consistency. Not surprisingly, Teamwork was best predicted by Hostility to Authority (r = -.18) and Cultural Tolerance (r = .14, both p < .05).

With respect to the attitudinal criteria, five RBI scales demonstrated validities in excess of r =.25 (all p < .05) for predicting Army Satisfaction, and six RBI scales did the same for predicting Perceived Fit. Among the best predictors of these two criteria were Achievement Orientation, Internal Locus of Control, Respect for Authority, Hostility to Authority, and Gratitude.

Prediction of Attrition Cognitions peaked in the high .20s, with the estimated validities of Fitness Motivation, Hostility to Authority, Internal Locus of Control, Respect for Authority, and Gratitude equal to or exceeding r = .20 (p < .05). The RBI Achievement Orientation and Respect for Authority scales were the best predictors of Career Intentions (r = .19 and r = .22, respectively, p < .05) and of Future Army Affect (r = .26 and r = .21, respectively, p < .05).

Table 9.2 also presents validity estimates corrected for criterion unreliability and range restriction due to selection on the AFQT. Among the performance criteria, the Hostility to Authority (r = -.33) and Cultural Tolerance (r = .24) scales were the strongest predictors of Teamwork. The Physical Fitness criterion was best predicted by Fitness Motivation (r = .34). The best predictors of the remaining three performance criteria included the Peer Leadership, Cognitive Flexibility, Achievement Orientation, Hostility to Authority, Self Esteem, Internal Locus of Control, and Army Identification scales, with median corrected validities ranging between .27 and .29.

With respect to the attitudinal criteria, Satisfaction with the Army was best predicted by Internal Locus of Control, Respect for Authority, Hostility to Authority, and Achievement Orientation (all r > .30). Perceived Fit was best predicted by Achievement Orientation and Respect for Authority (each r > .38). The same two scales were also the strongest predictors of Attrition Cognitions (r = -.21 and .23, respectively), and Future Army Affect (r = .28 and r = .23, respectively).

The Army Identification scale showed the highest pattern of correlations with the attitudinal criteria, as expected given the criterion contamination of this scale. Specifically, the RBI Army Identification was most closely related to the respondent's satisfaction with the Army

and the respondent's perceived fit with the Army. Excluding the Army Identification scale, the RBI scales of Achievement Orientation and Respect for Authority were the most consistent predictors of attitudinal criteria.

Incremental Validity Estimates

The uncorrected incremental validities of the RBI scales over and above AFQT for predicting the performance criteria are presented in Table 9.3. The RBI yielded the largest incremental validity over AFQT for the Achievement and Effort and Physical Fitness criteria. Twelve of the 14 RBI predictor scales demonstrated significant incremental validity for Achievement/Effort, with the Hostility to Authority, Achievement Orientation, and Army Identification showing *increases* in validities similar in magnitude (13 to 15 points) to the bivariate correlation between AFQT and the criterion. This is not too surprising given that the RBI is an assessment of individual motivation.

The RBI also showed potential for improving prediction of the Physical Fitness criterion over and above the AFQT. The modified Fitness Motivation RBI scale yielded a 30-point increase in validity over AFQT when predicting this criterion. Incremental validities ranging between 11 and 12 points were obtained for this criterion with the Peer Leadership, Achievement Orientation, Self-Esteem, and Army Identification scales.

With respect to the remaining three performance criteria, the Hostility to Authority and Cultural Tolerance scales demonstrated incremental validities of nine and seven points, respectively, for predicting Teamwork. The magnitudes of incremental validities for the General Technical Performance criterion were more modest, perhaps in part due to the strength of the AFQT's correlation with this criterion. Regardless, an increase of four to five points over the AFQT was still obtained from the RBI Peer Leadership, Fitness Motivation, Self-Esteem, and Army Identification scales. A similar pattern of results was obtained for the Expected Future Performance criterion.

Regarding the attitudinal criteria, it is not surprising that the RBI scales added significant incremental validity to the AFQT because the AFQT did not strongly predict these criteria. The strongest pattern of incremental validities was obtained for the Army Satisfaction and Perceived Fit criteria. Achievement Orientation, Internal Locus of Control, and Respect for Authority were the best predictors of Army Satisfaction (all $r > .26$, $p < .05$) and Perceived Fit (all $r > .32$, $p < .05$).

RBI scale incremental validities were not as high for the other attitudinal criteria. However, Fitness Motivation and Internal Locus of Control achieved the best incremental validity for Attrition Cognitions (both $r = .13$, $p < .05$). Respect for Authority had the highest incremental validity for Career Intentions ($r = .17$, $p < .05$), and Achievement Orientation demonstrated the best incremental validity for Future Army Affect ($r = .23$, $p < .05$)

Table 9.3 also presents corrected incremental validities for the RBI scales over AFQT. For the performance criteria, incremental validities as high as four points for Expected Future Performance, three points for General Technical Proficiency, 11 points for Achievement and Effort, 29 points for Physical Fitness, and 11 points for Teamwork were obtained with the RBI predictor scales. Of course, these results reflect the use of RBI scales individually, so they do not

reflect the potential for incremental validity above the AFQT when the RBI scales are used in combination.

Table 9.3. Incremental Validity Estimates for RBI Scales

	Performance Criteria					Attitudinal Criteria				
Scale	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Incremental Validity Estimates									
Peer Leadership	**.04**	**.06**	**.11**	.00	**.04**	.05	**.23**	**.02**	**.07**	**.15**
Cognitive Flexibility	.01	**.03**	.02	.02	.01	**.09**	**.19**	.02	**.06**	**.16**
Achievement Orientation	**.02**	**.13**	**.12**	**.03**	**.03**	**.27**	**.39**	**.07**	**.15**	**.23**
Fitness Motivation	**.04**	**.05**	**.30**	.00	**.05**	**.20**	**.27**	**.13**	**.09**	**.14**
Interpersonal Skills–Diplomacy	**.03**	**.03**	**.09**	.01	**.02**	**.17**	**.22**	**.05**	**.04**	**.07**
Stress Tolerance	**.02**	.02	**.09**	.01	**.02**	**.23**	**.17**	**.08**	**.05**	.05
Hostility to Authority	**.03**	**.15**	.01	**.09**	**.05**	**.29**	**.27**	**.09**	**.10**	**.08**
Self-Esteem	**.04**	**.10**	**.12**	.02	**.05**	**.12**	**.24**	**.07**	**.10**	**.12**
Cultural Tolerance	.01	**.09**	.00	**.07**	**.03**	**.13**	**.20**	**.06**	.04	**.14**
Internal Locus of Control	**.02**	**.10**	**.07**	.02	.01	**.30**	**.33**	**.13**	**.11**	**.13**
Army Identification	**.05**	**.14**	**.12**	.01	**.04**	**.54**	**.68**	**.35**	**.41**	**.45**
Respect for Authority	.00	**.05**	.02	.00	.00	**.29**	**.35**	**.10**	**.17**	**.19**
Narcissism	.00	.00	**.10**	.03	.00	.01	.07	.00	.01	**.10**
Gratitude	.01	**.07**	.00	.02	.01	**.25**	**.31**	**.16**	**.08**	**.11**
Lie Scale	.00	.02	.03	.01	.00	**.08**	**.11**	.02	.01	.02
	Corrected Incremental Validity Estimates									
Peer Leadership	.03	.04	.09	.00	.03	.00	.24	.01	.04	.13
Cognitive Flexibility	.00	.01	.00	.01	.00	.05	.19	.01	.03	.14
Achievement Orientation	.01	.09	.09	.03	.02	.27	.43	.05	.11	.22
Fitness Motivation	.02	.02	.29	.00	.04	.19	.29	.11	.05	.11
Interpersonal Skills–Diplomacy	.02	.01	.05	.00	.01	.16	.23	.03	.00	.03
Stress Tolerance	.01	.00	.05	.00	.01	.22	.17	.06	.02	.00
Hostility to Authority	.02	.11	.00	.11	.04	.29	.29	.07	.07	.04
Self-Esteem	.02	.07	.09	.01	.04	.10	.25	.05	.06	.10
Cultural Tolerance	.00	.06	.00	.07	.02	.11	.21	.04	.00	.12
Internal Locus of Control	.01	.07	.03	.01	.00	.31	.35	.11	.07	.11
Army Identification	.03	.11	.09	.00	.03	.56	.76	.35	.39	.46
Respect for Authority	.00	.03	.00	.00	.00	.29	.38	.08	.14	.17
Narcissism	.00	.00	.07	.02	.00	.00	.02	.00	.00	.06
Gratitude	.00	.04	.00	.01	.00	.25	.33	.14	.05	.08
Lie Scale	.00	.00	.00	.00	.00	.05	.10	.01	.00	.00

Note. n = 487 (AE criterion), N = 634-636 (all other Performance criteria), N = 611-631 (Attitudinal criteria). Cell values for the AFQT represent zero-order correlations between the AFQT and the given criterion (shown for reference). Uncorrected incremental estimates reflect the difference between the Multiple R obtained when regressing the criterion on both the given composite and AFQT versus the R obtained when regressing the criterion only on the AFQT. Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Statistically significant incremental validities are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

For the attitudinal criteria, incremental validities as high as 29 points were observed for Army Satisfaction, 43 points for Perceived Fit, 14 points for Attrition Cognitions, 14 points for Career Intentions, and 22 points for Future Army Affect. The RBI scales of Achievement Orientation, Internal Locus of Control, and Respect for Authority showed the best pattern of validities.

Subgroup Differences

The effect sizes of the RBI scales for gender and race are presented in Tables 9.4 and 9.5, respectively. There were no significant differences between female and male Soldiers for eight of the 14 RBI predictor scales. Females tended to score higher than males in Cognitive Flexibility, Achievement Orientation, Cultural Tolerance, and Respect for Authority. Females scored lower than males in Hostility to Authority and Fitness Motivation.

Table 9.4. RBI Scale Scores by Gender

		Male		Female	
Scale	d_{FM}	*M*	*SD*	*M*	*SD*
Peer Leadership	0.06	3.38	0.65	3.42	0.64
Cognitive Flexibility	**0.29**	3.40	0.72	3.61	0.62
Achievement Orientation	**0.50**	3.26	0.63	3.57	0.50
Fitness Motivation	**-0.36**	3.50	0.68	3.25	0.69
Interpersonal Skills–Diplomacy	0.10	3.40	0.80	3.48	0.82
Stress Tolerance	-0.22	2.88	0.51	2.77	0.48
Hostility to Authority	**-0.56**	2.80	0.67	2.42	0.59
Self-Esteem	-0.14	3.89	0.60	3.81	0.50
Cultural Tolerance	**0.42**	3.70	0.79	4.03	0.60
Internal Locus of Control	0.03	3.36	0.60	3.37	0.54
Army Identification	-0.17	3.08	0.83	2.94	0.80
Respect for Authority	**0.28**	3.31	0.68	3.50	0.59
Narcissism	-0.11	3.63	0.59	3.57	0.52
Gratitude	0.11	3.82	0.73	3.90	0.64
Lie Scale	-0.21	0.06	0.09	0.04	0.07

Note. n_{Male}= 589-614. n_{Female}= 73. d_{FM}= Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males/*SD* of males. Statistically significant effect sizes are bolded, p <.05 (two-tailed).

Hispanic Soldiers scored similarly to White Soldiers on most RBI scales, although Hispanics had relatively higher scores on the Cultural Tolerance scale. Furthermore, they also tended to trigger more faking items than Whites. Like Hispanics, Black Soldiers scored substantially higher than Whites in Cultural Tolerance. Blacks also scored higher in Achievement Orientation and Narcissism relative to Whites. The largest Black/White difference was seen in the Army Identification scale, with Black Soldiers scoring roughly one-half *SD* lower. It could be the case that Black Soldiers are more likely to enlist in the Army because of the opportunities for career training rather than because they are intrinsically interested in being a Soldier. This and other related hypotheses might be interesting topics for future research.

Table 9.5. RBI Scale Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
Scale	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Peer Leadership	0.09	-0.09	3.39	0.65	3.45	0.64	3.40	0.67	3.35	0.58
Cognitive Flexibility	0.02	-0.07	3.42	0.74	3.44	0.62	3.44	0.76	3.38	0.66
Achievement Orientation	**0.27**	0.07	3.25	0.64	3.42	0.56	3.24	0.62	3.29	0.66
Fitness Motivation	-0.04	**0.21**	3.48	0.68	3.45	0.72	3.45	0.70	3.60	0.58
Interpersonal Skills–Diplomacy	0.19	0.09	3.39	0.82	3.55	0.73	3.36	0.83	3.44	0.81
Stress Tolerance	-0.04	0.16	2.87	0.52	2.85	0.48	2.85	0.52	2.93	0.51
Hostility to Authority	-0.04	0.13	2.77	0.68	2.74	0.67	2.74	0.65	2.83	0.74
Self-Esteem	0.10	0.02	3.88	0.60	3.94	0.53	3.88	0.59	3.89	0.62
Cultural Tolerance	**0.30**	**0.39**	3.68	0.80	3.92	0.68	3.62	0.83	3.93	0.65
Internal Locus of Control	0.02	-0.04	3.35	0.62	3.36	0.50	3.35	0.62	3.32	0.60
Army Identification	**-0.54**	0.02	3.16	0.82	2.71	0.79	3.15	0.85	3.16	0.72
Respect for Authority	0.09	0.07	3.31	0.67	3.37	0.66	3.30	0.66	3.35	0.69
Narcissism	**0.51**	0.09	3.55	0.58	3.85	0.52	3.55	0.58	3.61	0.59
Gratitude	0.00	0.03	3.83	0.74	3.82	0.65	3.82	0.76	3.84	0.70
Lie Scale	0.17	**0.23**	0.05	0.09	0.07	0.09	0.05	0.09	0.07	0.09

Note. n_{White} = 476-494. n_{Black} = 124-129. $n_{White\text{-}Non\text{-}Hispanic}$ = 374-387. $n_{Hispanic}$=128-134. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites)/*SD* of Whites. Referent groups (e.g., White) are listed second in the effect size subscript. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Differential Prediction

Tests of slope and intercept differences by gender are presented in Table 9.6. With respect to slope differences, most comparisons revealed no significant differences, and the number of significant differences detected was near the Type 1 error rate. The exception was the Teamwork criterion where six of the 14 RBI predictor scales showed gender differences. Scale validities for females exceeded those for males each of these six cases, with differences in validity coefficients ranging between 16 points and 20 points. A larger number of intercept differences achieved statistical significance, particularly against the Achievement/Effort, Teamwork, Future Expected Performance, Attrition Cognitions, and Future Army Affect criteria. Use of a common regression line in these instances would result in the underprediction of females' performance.

Presented in Table 9.7 are tests of slope and intercept differences by race. Again, few slope differences were detected, and the frequency of the significant differences detected across all criteria was roughly what would be expected by chance. As with gender, many more race intercept differences were detected. Use of a common regression line would tend to overpredict the performance of blacks for the General Technical Performance, Achievement and Effort, Future Expected Performance, and Future Army Affect criteria. Moreover, a common regression line would under-predict the Attrition Cognitions of blacks.

Tests of slope and intercept differences by ethnic group are presented in Table 9.8. Once again, very few slope differences were detected. Intercept difference tests revealed that use of a common regression line would under-predict the performance of Hispanics for the Teamwork and Future Army Affect criteria.

Table 9.6. Differential Prediction Results for RBI Scales by Gender

	Performance Criteria																			
	GTP				AE				PF				TEAM				FXP			
Scale	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F
Peer Leadership	-.01	**.15**	**.20**	**.46**	**.25**	.12	**.17**	**.37**	-.10	.00	**.15**	.12	**.22**	**.18**	.02	**.29**	**.21**	.03	**.16**	.20
Cognitive Flexibility	-.06	.11	**.16**	**.32**	**.21**	.11	**.14**	**.30**	-.15	.15	.03	.18	**.16**	**.20**	.06	**.31**	**.17**	.09	**.12**	**.23**
Achievement Orientation	-.08	.08	**.15**	**.24**	**.17**	.06	**.24**	**.29**	-.21	.12	**.15**	**.23**	.15	.10	.07	.18	.13	.10	**.12**	**.21**
Fitness Motivation	.02	.01	**.17**	.20	**.28**	-.02	**.17**	.13	-.04	-.08	**.34**	**.21**	**.25**	.12	.00	.19	**.25**	-.02	**.18**	.16
Interpersonal Skills–Diplomacy	-.02	.08	**.14**	**.29**	**.25**	.03	**.12**	.18	-.12	.12	**.10**	**.25**	**.20**	**.18**	.01	**.31**	**.20**	.04	**.11**	.17
Stress Tolerance	.01	.01	**.17**	.18	**.28**	.07	**.12**	**.24**	-.07	.07	**.11**	.18	**.23**	.09	.06	.18	**.25**	.11	**.12**	**.27**
Hostility to Authority	-.06	.03	**-.21**	-.13	**.20**	.04	**-.30**	-.19	-.18	-.15	.03	-.14	.15	.00	**-.17**	-.13	.14	-.01	**-.17**	-.16
Self-Esteem	.03	.13	**.20**	**.36**	**.29**	.09	**.23**	**.34**	-.08	.03	**.14**	.14	**.25**	**.21**	.06	**.32**	**.24**	.08	**.19**	**.25**
Cultural Tolerance	-.12	**.21**	.05	**.36**	.13	**.23**	**.16**	**.47**	-.18	.19	.01	.18	.10	**.20**	**.11**	**.32**	.09	**.21**	**.09**	**.32**
Internal Locus of Control	-.02	.14	**.14**	**.36**	**.25**	.08	**.23**	**.35**	-.11	.11	**.09**	.20	**.20**	.09	.07	.18	**.20**	.14	.07	**.26**
Army Identification	.01	.06	**.18**	**.27**	**.28**	.03	**.27**	**.32**	-.11	-.11	**.16**	.02	**.24**	**.16**	.02	**.26**	**.23**	.05	**.14**	.21
Respect for Authority	-.03	.03	.02	.08	**.25**	-.03	**.14**	.08	-.09	-.16	.05	-.13	**.19**	.06	.01	.09	**.19**	.05	.03	.09
Narcissism	-.03	-.04	-.04	-.11	**.25**	.01	-.04	-.01	-.11	-.10	**.13**	.00	**.20**	-.03	**-.08**	-.10	**.20**	.00	-.05	-.05
Gratitude	-.03	.08	**.10**	**.22**	**.24**	.03	**.18**	.20	-.11	-.04	.02	-.03	**.19**	.09	.06	.17	**.20**	.04	**.08**	.11
Lie Scale	-.01	.05	-.04	.05	**.27**	.05	.07	.11	-.09	.06	.04	.08	**.24**	.13	.02	.16	**.24**	.14	.01	.16

Table 9.6. (Continued)

	Attitudinal Criteria																			
	ASat				AFit				ACog				CInt				FAA			
	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F	Int.	Slope	r_M	r_F
Peer Leadership	-.17	.10	.05	.19	-.06	.09	**.23**	**.30**	**.38**	-.20	**-.09**	**-.25**	-.02	**.32**	**.08**	**.35**	**-.26**	-.02	**.18**	.15
Cognitive Flexibility	-.18	-.03	**.10**	.06	-.08	-.03	**.19**	.13	**.42**	-.10	**-.12**	-.18	-.04	.09	**.07**	.13	**-.27**	-.11	**.19**	.06
Achievement Orientation	**-.31**	.09	**.29**	**.35**	**-.23**	.05	**.41**	**.37**	**.51**	-.14	**-.18**	**-.24**	-.23	**.34**	**.17**	**.39**	**-.37**	-.03	**.28**	**.21**
Fitness Motivation	-.12	-.05	**.22**	.16	.01	-.08	**.29**	.18	**.33**	.18	**-.24**	-.05	.06	.00	**.13**	.13	-.19	.02	**.15**	.19
Interpersonal Skills – Diplomacy	-.17	-.09	**.20**	.09	-.06	-.05	**.23**	.16	**.37**	.08	**-.15**	-.06	-.02	.18	.06	**.23**	**-.26**	.01	**.10**	.11
Stress Tolerance	-.11	.04	**.23**	**.28**	.02	.18	**.15**	**.35**	**.30**	-.12	**-.17**	**-.26**	.07	.22	.07	**.26**	-.22	.08	.06	.14
Hostility to Authority	**-.28**	.01	**-.32**	**-.29**	**-.23**	-.16	**-.27**	**-.40**	**.57**	.21	**.21**	**.36**	-.18	-.24	**-.12**	**-.30**	**-.28**	.05	**-.12**	-.06
Self-Esteem	-.14	.05	**.13**	.18	-.01	.01	**.24**	.20	**.32**	-.16	**-.15**	**-.25**	.05	.16	**.12**	**.23**	**-.25**	-.16	**.16**	-.02
Cultural Tolerance	-.20	-.05	**.17**	.08	-.19	.20	**.20**	**.32**	**.55**	-.30	**-.15**	**-.32**	-.24	**.56**	.04	**.41**	**-.38**	.14	**.18**	**.26**
Internal Locus of Control	-.16	-.06	**.33**	**.23**	-.05	-.03	**.34**	**.26**	**.36**	-.01	**-.26**	**-.22**	.01	.05	**.14**	.17	**-.24**	-.13	**.17**	.03
Army Identification	-.07	.05	**.55**	**.61**	.09	.12	**.68**	**.75**	**.26**	-.08	**-.46**	**-.48**	.13	.10	**.45**	**.52**	-.17	-.04	**.48**	**.45**
Respect for Authority	**-.22**	-.02	**.32**	**.28**	-.12	-.01	**.36**	**.29**	**.34**	**.35**	**-.24**	.10	-.06	.03	**.22**	**.22**	**-.28**	-.08	**.23**	.13
Narcissism	-.17	-.07	-.01	-.10	-.06	**-.22**	**.10**	-.15	**.36**	.00	.01	.01	.00	-.05	.03	-.01	**-.25**	-.13	**.13**	.00
Gratitude	-.17	-.15	**.29**	.09	-.06	-.10	**.33**	.17	**.37**	.16	**-.30**	-.11	.00	-.13	**.13**	.01	**-.24**	-.23	**.16**	-.09
Lie Scale	-.15	-.01	**.10**	.06	-.08	-.23	**.14**	-.11	**.41**	.30	**-.08**	.16	-.01	-.12	.04	-.05	**-.25**	-.05	.04	-.01

Note. n_{Male} = 448-592. n_{Female} = 57-71. Int. = Unstandardized regression weight for gender (0 = male, 1 = female). Slope = Unstandardized regression weight for the RBI by- gender interaction term. This weight reflects the difference between unstandardized regression weights for males and females on the given RBI scale ($b_{RBI,females}$ - $b_{RBI,males}$) based on the full regression model. r_M = Correlation between the given RBI scale and the given criterion for males. r_F = Correlation between the given RBI scale and the given criterion for females. Statistically significant regression weights are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 9.7. Differential Prediction Results for RBI Scales by Race

	Performance Criteria																			
	GTP				AE				PF				TEAM				FXP			
Scale	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B
Peer Leadership	**-.26**	-.03	**.25**	**.21**	**-.16**	-.05	**.23**	.10	.01	.05	**.15**	**.21**	-.04	.02	.06	.08	**-.16**	-.02	**.19**	**.19**
Cognitive Flexibility	**-.25**	.03	**.18**	**.21**	**-.16**	.00	**.18**	.13	.02	.03	.04	.07	-.03	.07	**.09**	**.16**	**-.15**	.04	**.14**	**.21**
Achievement Orientation	**-.28**	.01	**.18**	**.20**	**-.18**	-.07	**.33**	.16	-.03	.06	**.15**	**.21**	-.05	-.02	**.11**	.07	**-.18**	-.01	**.17**	**.18**
Fitness Motivation	**-.25**	-.07	**.19**	.08	**-.16**	-.06	**.15**	.04	.03	-.05	**.34**	**.30**	-.03	-.04	.03	-.05	**-.15**	-.05	**.17**	.14
Interpersonal Skills–Diplomacy	**-.26**	-.02	**.18**	.14	**-.16**	-.09	**.20**	.01	.01	-.03	**.16**	.10	-.04	.03	.05	.09	**-.15**	-.09	**.15**	.00
Stress Tolerance	**-.24**	.05	**.14**	**.25**	**-.16**	-.05	**.14**	.04	.03	.03	**.12**	.14	-.03	.09	.04	**.16**	**-.15**	.04	**.10**	**.18**
Hostility to Authority	**-.25**	-.03	**-.20**	**-.30**	**-.17**	-.05	**-.29**	**-.35**	.02	.08	-.01	.10	-.04	-.06	**-.18**	**-.28**	**-.16**	-.01	**-.19**	**-.25**
Self-Esteem	**-.26**	.03	**.22**	**.29**	**-.16**	-.02	**.26**	**.20**	.01	-.02	**.18**	.14	-.04	.05	**.09**	.14	**-.17**	.06	**.20**	**.33**
Cultural Tolerance	**-.28**	.01	**.11**	.12	**-.20**	-.04	**.25**	.14	.03	-.17	.06	-.15	-.07	-.03	**.17**	.09	**-.17**	-.09	**.16**	.02
Internal Locus of Control	**-.25**	.05	**.15**	**.23**	**-.16**	.01	**.25**	**.19**	.02	-.12	**.13**	-.03	-.04	.10	.06	**.19**	**-.15**	-.03	**.10**	.05
Army Identification	**-.21**	-.01	**.16**	.15	-.07	.00	**.27**	**.22**	.08	-.03	**.17**	.12	-.02	.01	.04	.05	-.11	-.01	**.13**	.14
Respect for Authority	**-.25**	-.02	.05	.01	**-.16**	**-.16**	**.24**	-.05	.02	.02	.02	.05	-.03	-.09	.07	-.08	**-.15**	.00	.05	.06
Narcissism	**-.25**	.01	-.01	.01	**-.15**	-.02	.01	-.03	-.06	.06	**.12**	**.18**	.00	.00	**-.10**	-.08	**-.17**	.08	-.05	.07
Gratitude	**-.25**	.00	**.11**	.11	**-.16**	-.08	**.24**	.06	.02	-.08	.05	-.06	-.03	-.07	**.11**	-.02	**-.15**	-.05	**.11**	.03
Lie Scale	**-.25**	.05	-.05	.04	**-.16**	.00	.05	.05	.02	-.05	.05	-.02	-.04	.05	.00	.08	**-.16**	.02	.01	.04

Table 9.7. (Continued)

	Attitudinal Criteria																			
	ASat				AFit				ACog				CInt				FAA			
	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B	Int.	Slope	r_W	r_B
Peer Leadership	-.09	-.02	**.08**	.05	-.13	-.02	**.25**	**.21**	**.37**	-.07	**-.10**	**-.17**	.08	.05	**.10**	.14	**-.21**	.10	**.17**	**.26**
Cognitive Flexibility	-.08	.02	**.09**	.10	-.12	-.05	**.18**	.10	**.36**	-.07	**-.10**	-.14	.09	-.09	**.09**	.00	-.19	-.05	**.19**	.11
Achievement Orientation	-.12	-.14	**.32**	.12	**-.18**	-.12	**.43**	**.24**	**.41**	.00	**-.19**	**-.18**	.07	-.23	**.22**	.01	**-.24**	-.05	**.28**	**.19**
Fitness Motivation	-.08	**-.17**	**.26**	.04	-.11	**-.21**	**.31**	.05	**.35**	.08	**-.22**	-.14	.09	-.18	**.16**	-.01	-.17	-.12	**.18**	.06
Interpersonal Skills – Diplomacy	-.10	-.08	**.22**	.09	-.13	-.15	**.26**	.05	**.39**	-.01	**-.16**	**-.16**	.11	-.23	**.11**	-.09	**-.20**	.01	**.11**	.11
Stress Tolerance	-.08	-.01	**.24**	**.21**	-.12	-.07	**.19**	.09	**.35**	-.01	**-.19**	**-.19**	.09	-.20	**.10**	-.07	-.18	-.12	**.11**	-.02
Hostility to Authority	-.08	.05	**-.30**	**-.24**	-.12	.03	**-.27**	**-.23**	**.36**	.07	**.18**	**.27**	.10	.21	**-.17**	.02	-.18	.11	**-.11**	.00
Self-Esteem	-.08	**-.17**	**.18**	-.04	-.12	**-.19**	**.27**	.03	**.37**	.01	**-.17**	**-.16**	.11	**-.30**	**.17**	-.09	-.17	**-.23**	**.20**	-.05
Cultural Tolerance	-.10	-.13	**.18**	.00	-.17	-.12	**.24**	.08	**.44**	-.07	**-.15**	**-.19**	.12	-.18	**.09**	-.06	**-.25**	-.10	**.22**	.10
Internal Locus of Control	-.08	-.07	**.34**	**.20**	-.12	-.10	**.35**	**.17**	**.35**	.06	**-.26**	**-.17**	.09	**-.33**	**.17**	-.11	-.18	-.07	**.17**	.08
Army Identification	**.16**	-.02	**.57**	**.53**	**.20**	.01	**.70**	**.67**	.10	.04	**-.47**	**-.44**	**.38**	-.06	**.49**	**.42**	.11	.03	**.49**	**.49**
Respect for Authority	-.10	.04	**.30**	**.35**	-.14	-.03	**.35**	**.32**	**.37**	-.01	**-.20**	**-.23**	.08	.00	**.21**	**.21**	**-.20**	-.07	**.23**	.15
Narcissism	-.07	-.04	.01	-.03	-.13	-.08	**.10**	.00	**.37**	.01	-.03	-.02	.07	.00	.03	.02	-.19	-.14	**.15**	.00
Gratitude	-.08	-.11	**.31**	.14	-.12	-.06	**.33**	**.21**	**.35**	.17	**-.31**	-.12	.09	**-.30**	**.16**	-.10	**-.19**	.06	**.13**	**.16**
Lie Scale	-.08	**-.17**	**.15**	-.06	-.12	**-.24**	**.19**	-.11	**.35**	.19	**-.09**	.09	.09	-.15	**.08**	-.05	-.17	**-.23**	**.10**	-.14

Note. n_{White} = 359-475. n_{Black} = 91-120. Int. = Unstandardized regression weight for race (0 = White, 1 = Black). Slope = Unstandardized regression weight for the RBI by race interaction term. This weight reflects the difference between unstandardized regression weight for Whites and Blacks on the given RBI scale ($b_{RBI,Blacks}$ - $b_{RBI,Whites}$) based on the full regression model. r_W = Correlation between the given RBI scale and the given criterion for Whites. r_B = Correlation between the given RBI scale and the given criterion for Blacks. Statistically significant regression weights are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 9.8. Differential Prediction Results for RBI Scales by Ethnic Group

	Performance Criteria																			
	GTP				AE				PF				TEAM				FXP			
Scale	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H
Peer Leadership	-.03	-.03	**.26**	**.21**	.09	.02	**.23**	**.26**	.12	-.02	**.16**	.12	**.18**	.07	.03	**.15**	.10	.00	**.20**	**.21**
Cognitive Flexibility	-.03	.05	**.16**	**.26**	.09	-.01	**.20**	.16	.12	.06	.02	.09	**.18**	**.12**	.05	**.27**	.10	.04	**.13**	**.21**
Achievement Orientation	-.06	-.04	**.19**	**.15**	.06	-.07	**.35**	**.26**	.10	.01	**.13**	**.16**	**.17**	.05	.08	**.20**	.07	-.07	**.18**	.12
Fitness Motivation	-.07	.00	**.20**	**.20**	.05	.05	**.14**	**.23**	.04	.07	**.32**	**.36**	**.16**	.03	.01	.07	.05	.04	**.16**	**.22**
Interpersonal Skills–Diplomacy	-.05	-.07	**.21**	.09	.09	**-.20**	**.25**	-.16	.11	-.14	**.19**	-.01	**.17**	-.03	.04	-.01	.08	-.09	**.17**	.04
Stress Tolerance	-.06	-.02	**.16**	.14	.08	-.06	**.13**	.02	.11	-.09	**.12**	.01	**.17**	-.02	.03	-.01	.08	-.01	**.10**	.10
Hostility to Authority	-.04	.05	**-.21**	**-.15**	.11	-.02	**-.29**	**-.37**	.12	-.04	.02	-.04	**.18**	.10	**-.21**	-.06	.09	**.15**	**-.23**	-.02
Self-Esteem	-.05	-.01	**.21**	**.25**	.07	-.01	**.26**	**.28**	.11	**-.15**	**.21**	.01	**.17**	.08	.04	**.22**	.08	.01	**.19**	**.26**
Cultural Tolerance	-.07	-.04	**.12**	.05	.03	.01	**.24**	**.24**	.11	-.07	.06	-.03	**.15**	.00	**.14**	.12	.05	-.05	**.16**	.08
Internal Locus of Control	-.04	.04	**.12**	**.22**	.09	.06	**.22**	**.36**	.11	-.04	**.13**	.07	**.17**	-.01	.05	.04	.09	.00	.08	.10
Army Identification	-.05	.02	**.16**	**.19**	.08	.03	**.26**	**.29**	.09	.12	**.13**	**.26**	**.18**	-.10	.07	-.10	.08	.02	**.13**	**.16**
Respect for Authority	-.05	-.03	.05	-.01	.08	-.10	**.26**	.08	.13	.03	.00	.04	**.17**	-.02	.06	.04	.08	.00	.04	.05
Narcissism	-.05	.01	-.01	.00	.09	.03	-.02	.04	.11	-.01	**.12**	.12	**.18**	.07	**-.11**	.01	.09	.06	-.06	.04
Gratitude	-.05	.01	**.10**	.13	.08	**-.13**	**.27**	.01	.11	-.04	.05	-.01	**.17**	-.04	**.11**	.04	.09	-.07	**.11**	.01
Lie Scale	-.04	-.06	-.02	-.14	.07	.03	.05	.12	.12	-.13	**.10**	-.08	**.17**	.00	.00	.01	.08	.02	-.01	.03

Table 9.8. (Continued)

	Attitudinal Criteria																			
	ASat				AFit				ACog				CInt				FAA			
	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H	Int.	Slope	r_W	r_H
Peer Leadership	.12	-.08	**.09**	-.01	.12	.01	**.24**	**.24**	.04	.05	**-.11**	-.06	.05	.10	**.09**	**.18**	**.25**	.00	**.17**	**.18**
Cognitive Flexibility	.12	.05	.07	.12	.11	.07	**.17**	**.25**	.04	-.13	-.08	**-.20**	.04	.03	.08	.10	**.25**	.03	**.18**	**.22**
Achievement Orientation	.09	-.13	**.35**	**.20**	.07	-.09	**.44**	**.37**	.07	**.19**	**-.24**	-.05	.02	-.18	**.25**	.10	**.22**	-.09	**.29**	**.25**
Fitness Motivation	.07	.03	**.25**	**.25**	.05	-.01	**.31**	**.27**	.14	**-.23**	**-.19**	**-.37**	-.01	.05	**.15**	**.18**	**.20**	.05	**.17**	**.22**
Interpersonal Skills – Diplomacy	.10	-.11	**.25**	.10	.09	-.10	**.30**	**.17**	.06	.09	**-.18**	-.09	.03	-.14	**.14**	.01	**.23**	-.05	**.13**	.08
Stress Tolerance	.09	-.09	**.26**	**.15**	.08	-.08	**.22**	.13	.07	.13	**-.23**	-.09	.01	.00	**.11**	.12	**.23**	-.08	**.12**	.04
Hostility to Authority	.14	.05	**-.30**	**-.28**	.13	.00	**-.27**	**-.32**	.03	-.09	**.21**	.14	.06	.19	**-.19**	-.04	**.25**	.00	**-.11**	**-.16**
Self-Esteem	.11	-.03	**.18**	**.16**	.10	-.01	**.27**	**.29**	.05	-.05	**-.16**	**-.23**	.03	.03	**.16**	**.21**	**.23**	.07	**.16**	**.30**
Cultural Tolerance	.10	-.15	**.21**	.01	.02	-.02	**.25**	**.20**	.08	.13	**-.19**	-.04	.04	-.25	**.12**	-.10	.19	-.17	**.24**	.05
Internal Locus of Control	.12	.01	**.33**	**.34**	.11	.05	**.33**	**.42**	.04	-.02	**-.26**	**-.29**	.04	-.04	**.18**	**.15**	**.24**	.09	**.14**	**.27**
Army Identification	.09	.04	**.58**	**.53**	.08	.03	**.70**	**.66**	.07	.00	**-.48**	**-.41**	.04	-.13	**.52**	**.36**	**.23**	.04	**.49**	**.53**
Respect for Authority	.12	-.03	**.30**	**.29**	.10	-.04	**.35**	**.35**	.05	-.01	**-.19**	**-.22**	.03	-.04	**.20**	**.19**	**.23**	.04	**.21**	**.32**
Narcissism	.12	.02	-.01	.02	.10	.03	**.09**	.14	.05	-.03	-.02	-.06	.03	-.02	.03	.02	**.23**	.01	**.14**	**.19**
Gratitude	.11	-.09	**.33**	**.19**	.09	-.02	**.33**	**.30**	.06	.11	**-.33**	**-.19**	.03	-.06	**.15**	.09	**.23**	.14	**.09**	**.27**
Lie Scale	.09	.00	**.13**	.13	.07	-.05	**.18**	.14	.07	.07	**-.12**	-.06	.02	-.02	.07	.06	**.21**	.09	.06	**.20**

Note. $n_{\text{White non-Hispanic}}$ = 284-371. n_{Hispanic} = 92-129. Int. = Unstandardized regression weight for ethnic group (0 = White non-Hispanic, 1 = Hispanic). Slope = Unstandardized regression weight for the RBI by ethnic group interaction term. This weight reflects the difference between unstandardized regression weight for White non-Hispanics and Hispanics on the given RBI scale ($b_{\text{RBI,Hispanics}}$ - $b_{\text{RBI,White non-Hispanics}}$) based on the full regression model. r_W = Correlation between the given RBI scale and the given criterion for White Non-Hispanics. r_H = Correlation between the given RBI scale and the given criterion for Hispanics. Statistically significant regression weights are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Discussion

Most of the RBI scales significantly predicted multiple indices of job performance and attitudinal criteria and also added incremental validity to AFQT for predicting both sets of outcomes. The magnitude of these statistics suggests that the RBI has good potential to augment the ASVAB in the enlisted accessions process. Thus, the results reported herein suggest that the RBI can help improve not only the prediction of enlisted job performance, but also attitudes that relate to attrition.

In addition to its usefulness as a predictor of important Army outcomes, the RBI would be relatively easy to implement in the accessions process. The test is short, requiring no more than 30 minutes to complete. It is easy to read and understand because the items are targeted at the third-grade reading level. The RBI uses multiple-choice questions along with an objective scoring key, making it easy to administer and score the test instantaneously on the web. Moreover, the format of the RBI makes it easy to develop parallel forms of the test and to update the test with scales measuring new predictor constructs. Finally, the modular nature of the test makes it simple to tailor the test for use in different settings.

Despite the evidence of promise for operational use, future research on the RBI is warranted. As noted earlier, the concurrent validation design produced artifactual criterion-related contamination for two RBI scales (Fitness Motivation and Army Identification). A longitudinal design is needed to determine whether these scales demonstrate good predictive validity with respect to attrition and job performance. As noted earlier, the prospect of predictive validity is suggested by the Select21 longitudinal attrition analyses indicating that both scales significantly predicted early Soldier attrition (Putka & Bradley, 2006). It would be useful for this longitudinal research to assess the test-retest reliability of the scales, a feature impossible to assess fully in a concurrent validation design. Additionally, because the RBI is a self-report measure, it is important to collect predictive validation data when the test is administered under operational conditions, where the motivation to fake is high (and where the referent used to respond to RBI items is different (i.e., pre-Army behavior vs. in-Army behavior). Other research suggests that biodata tests still are able to predict important outcomes (e.g., attrition, subsequent performance) when used in the selection process where the incentive to fake is high (Kilcullen, Goodwin, Chen, Wisecarver, & Sanders, in press). Although the operational use of the RBI may require the re-computation of cut-off scores to reflect some elevation of scores under these conditions, use of the lie-adjusted RBI scales will at least partially offset these elevations and help preserve the validity of the scales.

Given the relatively strong estimated validities obtained in this research, future research might look at the possibility of expanding the RBI to measure other important predictor constructs. In Select21, attitudes were one type of criterion measure, but in a longitudinal investigation, initial attitudes regarding the Army could serve as important predictor measures. In this light, it is interesting to note that the seven-item RBI Army Identification scale demonstrated strong convergent validity with affective commitment as indicated by the subject's perceived fit with the Army. This suggests that it may be possible to measure important non-cognitive predictors more efficiently and effectively by creating new RBI scales rather than administering large batteries of surveys and tests. An added benefit could be the capability of adjusting these scores to at least partially offset the effects of faking.

CHAPTER 10: WORK PREFERENCES SURVEY

Dan J. Putka and Chad H. Van Iddekinge
HumRRO

Overview

Personnel selection measures are typically designed to assess the knowledge, skills, and attributes (KSAs) critical to performance in the job of interest (Schmitt & Chan, 1998). Although important, job performance is not the only criterion of concern for most organizations. For example, the U.S. Army is interested in reducing attrition and increasing re-enlistment through personnel selection and classification. Traditional KSA-based selection measures, however, are seldom designed to predict both performance and alternative criteria such as attrition. In recent years, personnel researchers have turned to measures of person-environment (P-E) fit to predict criteria other than job performance (Ployhart, Schneider, & Schmitt, 2006). As part of the Select21 project, several P-E fit predictor measures were developed (Van Iddekinge, Putka, & Sager, 2005). In this chapter, we describe validation results for a vocational interests-based P-E fit predictor measure, the Work Preferences Survey.[26]

Instrument Description

The Work Preferences Survey (WPS) is a 72-item survey designed to assess a respondent's standing on Holland's (1985) RIASEC interest dimensions. According to Holland, vocational interests are expressions of personality that can be used to categorize individuals and work environments into six types: realistic, investigative, artistic, social, enterprising, and conventional (RIASEC). Holland's model has been widely validated and is the prevailing taxonomy in vocational psychology (Barrick, Mount, & Gupta, 2003).

The WPS contains three types of items. One type measures interests in work activities (e.g., "A job that requires me to teach others"), another measures interests in work environments (e.g., "A job that requires me to work outdoors"), and the final type measures interests in learning opportunities (e.g., "A job in which I can learn how to lead others"). Each item is designed to measure one of the six RIASEC dimensions. Respondents are asked to rate each item on a Likert-type scale with anchors that range from *extremely important to have in my ideal job* (1) to *extremely unimportant to have in my ideal job* (5). Item development was based on a thorough review of existing interest inventories and source materials from the vocational interest literature. Complete details on the development of the WPS are presented in Van Iddekinge, Putka et al. (2005).

The WPS produces six scale scores (one corresponding to each of the six RIASEC dimensions), and 14 facet scores (which represent components underlying the six RIASEC

[26] Soldiers participating in the Select21 concurrent validation effort were actually administered two vocational interest measures. In addition to the WPS, Soldiers completed the Department of Defense's Career Exploration Program Interests Inventory (CEP-II; Wall, Wise, & Baker, 1996). Like the WPS, the CEP-II was designed to measure a respondent's standing on Holland's (1985) RIASEC interest dimensions. In this chapter, we focus on evaluating the validity of the WPS, and we used CEP-II data only to examine construct-validity evidence for the WPS.

dimensions). Items for each scale/facet are averaged to create a total score for that scale/facet. Total scores on each facet/scale can range from one to five.

Method

Sample

A total of 784 Soldiers completed the WPS during the concurrent validation data collections (Wave 1 = 603, Wave 2 = 181). We eliminated the responses of 18 Soldiers who test administrators flagged as having questionable WPS data or that had exhibited extremely unlikely patterns of responding. Thus, the final analysis sample comprised 766 Soldiers (Wave 1 = 586, Wave 2 = 180).

Validation Strategy

A key element of any measure of P-E fit is how "environment-side" data (e.g., the extent to which the Army supports each of the RIASEC interests) are assessed and used in subsequent validation efforts (Kristof, 1996). The WPS, like other Select21 measures, is an assessment of person attributes (in this case, vocational interests) and does not reflect the extent to which a person's interests are supported by the Army environment. In earlier Select21 data collections, 107 Army NCOs completed the Army Environment Survey (AES), a measure designed to assess the degree to which the Army environment supports each of the RIASEC dimensions for first-term Soldiers. The development, administration, and psychometric properties of the AES were fully described in Van Iddekinge, Putka et al. (2005). We used mean NCO ratings from the AES on each RIASEC dimension as the environment-side "profile" when validating the WPS against the Select21 criteria.[27] Taken together, data from the WPS and AES can be combined to form an indirect, objective measure of P-E (Army) fit (Kristof-Brown, Zimmerman, & Johnson, 2005).

Although scoring the WPS and AES is straightforward, assessing the relationship between interests-related content and criterion measures has been a point of debate in the P-E fit and vocational counseling literature for decades (e.g., Camp & Chartrand, 1992; Edwards, 1991; Kristof, 1996; Kulka, 1979; Putka, 2005; Tinsley, 2000). Given this uncertainty, the question of how to best combine person (WPS) and environment (AES) data to predict various Select21 criteria (e.g., performance, satisfaction, career intentions) is an open one and, as such, so is the most appropriate strategy for estimating the criterion-related validity of such measures.

Given that several different methods exist in the P-E fit literature for evaluating relations between predictor content and criteria (see Edwards, 1991 for a review), we used this analysis as an opportunity to pit different methods against each other—something that has rarely been done in the fit literature. This method was the only defensible strategy since past research does not indicate the approach that is best for the Army to adopt. Given the above considerations, we constructed four types of WPS composites that we validated against the Select21 criteria: (a) traditional profile

[27] In previous Select21 data collections, a far smaller group of NCOs ($N = 6$) completed a future-oriented version of the AES—the Future Army Environment Survey (FAES). Although we initially considered creating fit measures based on comparison of the WPS and FAES, preliminary analyses suggested that the results we would achieve using the FAES would be very similar to those achieved using the AES (which is based on a far larger sample of NCOs). Thus, the AES served as the sole source of environment-side data.

similarity indexes (i.e., fit indexes), (b) regression weighted composites, (c) subjectively weighted composites, and (d) unit weighted composites. We discuss each of these in turn.

Traditional Profile Similarity Indexes

The first type of composite we constructed assesses the similarity (or dissimilarity) between a Soldier's interest profile on the WPS (at the scale-level) and the mean interest profile provided by NCOs on the AES. Such profile similarity indexes (PSIs) are the most common way person and environment data are combined in the vocational counseling and P-E fit literatures (Kristof, 1996). Two commonly used PSIs were calculated, and their criterion-related validities were estimated. The first index, D^2, reflects the sum of the squared differences between a Soldier's score on each RIASEC dimension and the mean SME score on each dimension. As Cronbach and Gleser (1953) noted, D^2 reflects differences in elevation (mean differences), scatter (variability differences), and shape (rank order differences) between a Soldier's WPS profile and NCOs' mean AES profile. The higher the D^2, the less similar a Soldier's profile is to the Army's profile. As a point of reference, if WPS scale scores for a given Soldier differed from each of the corresponding AES scores by .50, 1.0, and 2.0 scale points, the resulting D^2 values would be 1.5, 6.0, and 24.0, respectively.

The second profile similarity index we calculated was the correlation (Pearson *r*) between a Soldier's WPS profile and NCOs' mean AES profile. Unlike D^2, the correlation only assesses the similarity between profiles in terms of shape; it does not consider differences in elevation or scatter. Also, unlike D^2, higher Pearson *r* values indicate greater similarity.

Regression Weighted Composites

Although the fit indexes described above are useful for describing similarity of profiles, and are by far the most common strategy used for combining P-E data in the literature, past research has indicated that they put unrealistic and methodologically problematic constraints on person-environment-criterion (P-E-C) relations (e.g., Cable & Edwards, 2004; Edwards, 1991, 1993; Tinsley, 2000). For example, research has illustrated how using such fit indexes can limit the potential predictive validity of fit data (e.g., Edwards, 1993; Edwards & Parry, 1993). In light of such problems, many researchers have suggested viewing the constraints imposed by fit indexes on P-E-C relations as hypotheses to be tested using regression models (Cronbach, 1958; Edwards, 1993; Hesketh & Gardner, 1993; Tinsley, 2000). The most well known strategy for doing this is to use polynomial regression (Edwards, 1991, 1993). Using the predicted criterion score resulting from a polynomial regression model as a "predictor" has two distinct advantages over using a simple fit index as a "predictor" when estimating the criterion-related validity of P-E fit measures. First, it is advantageous from a theoretical perspective because it allows researchers to assess the viability of the constraints imposed on P-E-C relations by fit indexes and to better understand relations between individual profile elements (e.g., the RIASEC dimensions) and the criterion. Second, from a practical perspective, it allows researchers to free the aforementioned constraints and, in turn, fully realize the predictive validity of their person and environment data (see Edwards, 1993, and Putka, 2005, for illustrations of how fit indexes may constrain predictor-criterion relations).

Although a polynomial regression approach has benefits over fit indexes for predicting criteria, the approach has limited utility for Select21. Specifically, the approach is most applicable in situations where there is variation in environment-side data across persons in the validation sample. This situation was not the case in the present research, as the vocational interests profile for each Soldier was compared to a single environment profile, which was that of the Army in general. Putka (2005) illustrated how use of polynomial regression in such a situation is potentially problematic because it essentially excludes environment-side data from the modeling process. For this reason, Putka (2005) provided an extension of the regression-based approach to P-E fit (based on spline regression) designed to deal with this situation (i.e., by incorporating NCOs' mean AES data into the prediction model even though it is a constant across Soldiers). We used this approach to create regression weighted WPS composites for this validation effort. One regression weighted composite was constructed for each Select21 criterion (i.e., we attempted to create optimal composites for each criterion).[28]

Although the primary intent behind developing the WPS was to predict non-performance criteria (e.g., attrition and its attitudinal precursors), we also examined its validity for predicting the Select21 performance criteria. Researchers have begun to make the case that vocational interest and work value measures (cf. Chapter 11), which have not traditionally been used in selection contexts, may predict the "will-do" components of performance (Hogan & Hogan, 1996; Quintela, 2003). The rationale behind this idea is that, compared to traditional trait-based measures of personality (e.g., Big Five personality inventories), constructs such as interests and work values are more proximal to work motivation, a primary determinant of job performance (Campbell, McCloy, Oppler, & Sager, 1993). Specifically, interests and values are *directional* in nature (i.e., an expressed liking or preference to engage in some activity). Motivation has often been defined in terms of three elements, *direction* (choice to expend effort on some activity), *amplitude* (choice of level of effort to expend), and *duration* (choice to persist with that effort). Thus, a measure of interests such as the WPS can be looked at as a fairly proximal measure of *direction*, one of the key elements of motivation.[29] As such, we hypothesized that the WPS would predict Select21 performance composites that reflect will-do performance components (most notably the Achievement and Effort performance composite).

Subjective and Unit Weighted Composites

While the regression-based approaches to estimating the validity of P-E fit measures have some clear advantages over profile similarity indexes, a drawback of regression-based approaches is that their solutions may tend to be sample specific. That is, regression weights are optimized based on the sample in which the prediction model is estimated. As such, the multiple correlations (*R*) associated with such models capitalize upon chance, and may shrink upon cross-validation, particularly when they involve many predictor variables and higher order terms (Cattin, 1980).

Given this possibility, we also constructed subjectively weighted and unit weighted composites of WPS scales/facets for each Select21 criterion. These composites were constructed

[28] A regression weighted composite was not constructed for the Teamwork performance criterion due to its low reliability (see Chapter 5).

[29] We hypothesize that interests and values would be most proximal to the direction component of motivation, but acknowledge that Big Five facets and factors, such as the Achievement Striving facet of Conscientiousness, may be more proximal to the amplitude and duration components of motivation.

as follows. Once the regression weighted composite targeting a given criterion was formed, we calculated zero-order correlations between the given criterion and each WPS scale/facet that entered the final model for that criterion.[30] Only those WPS scales/facets with statistically significant estimated validities were included in the subjectively weighted and unit weighted composites for that criterion. For the unit weighted composites, all scales that entered the composite were given a weight of +1 or -1 (depending on the direction of their criterion-related validity). For the subjectively weighted composites, the majority of scales were also given a weight of +1 or -1, but in some cases, one of the scales/facets in the composite was given a weight of 2.0 or 0.5 (based on a large discrepancy between its criterion-related validity and the validity of other scales/facets in the composite).

It is important to note that although the subjectively weighted and unit weighted composites were not based on regression weights, their content reflects WPS scales/facets identified through the regression modeling described above. Therefore, the criterion-related validity of these composites will also likely shrink upon cross-validation, though we would expect the extent of shrinkage to be smaller than for the regression weighted composites.

Another key difference between the subjective/unit weighted composites and the regression weighted composites was that when forming the latter composites, data from the AES were considered in the modeling process. That is, when modeling a criterion called for inclusion of AES data in the prediction equation (e.g., via a spline adjustment term, or absolute difference term), they were included. In the case of the subjective/unit weighted composites, no AES data were included.

Although this process seems contrary to the point of constructing and validating measures of P-E fit, failure to consider the possibility that person-side data alone (i.e., only WPS data) may be sufficient to predict a given criterion has been a point of criticism in the literature (Tinsley, 2000). This possibility is most easily seen at the scale level. For example, if the Army environment provides either a very high or very low level of support for a given interest (e.g., Artistic interests), then Soldiers' scores on such an interest dimension would likely have a simple linear relation with the target criteria because "misfit" occurs in one direction only. In other words, incorporating AES data into the prediction equation through the addition of spline adjustment terms, or by using the WPS-AES absolute difference score as the predictor, would not increment prediction of the criterion (Putka, 2005). This notion is consistent with individual differences theory that has been the basis of personnel selection research since its inception, where non-linear relations between predictors and criteria are rarely found.

Cross-Validation

The various approaches to forming the WPS composites differ in terms of the degree to which their content and weighting are based on the sample data. As such, the criterion-related validity estimates for some of these composites may reflect capitalization on chance more than others. For example, the content of the profile similarity indexes (in terms of which WPS scales are included) is not at all dependent on the sample data, and as such, shrinkage is not an issue for these types of "composites." On the other hand, the content and weighting of the regression

[30] Appendix I of Knapp et al. (2005) describes how the regression composites were formed (see also Putka, 2005).

weighted composites were based entirely on the data. Not only were the data used to identify the proper functional form for the relation between each WPS scale and criterion for these composites, but the data were also used to determine how to weight the surviving WPS scales for forming a cross-scale composite. As such, we would expect criterion-related validity estimates for the regression composites to shrink upon cross-validation. Although the weights for the subjective and unit composites were not derived from regression analyses, their content partially reflects the regression results, and as such, the criterion-related validity of these composites would be expected to shrink to some extent.

Given that the construction of all "weighted" WPS composites was at least partially based on the data, it is desirable to have adjusted validity estimates that account for potential shrinkage. Under typical circumstances, the preferred approach would be to apply a shrinkage formula to the criterion-related validity estimate obtained in the full sample (Cattin, 1980). However, two issues make application of such formulae challenging in this case: (a) the multiple steps in the process of forming the regression weighted composites noted above, and (b) the partial dependence of the subjectively and unit weighted composites on the regression results. Thus, we adopted an alternative strategy for cross-validation.

As described in Chapter 2, concurrent validation data were collected in two waves. Upon completing the first wave of data collections, we constructed a set of WPS composites and presented findings to the Select21 Scientific Review Panel (SRP) in January of 2006. Upon meeting with the SRP, discussions ensued among project staff regarding how best to use the Wave 2 and full sample data for purposes of estimating the criterion-related validity of WPS content in light of the work that had already been done. On the one hand, there was a strong preference that the WPS composites be based on the full sample of data, yet doing so would create problems for cross-validating the resulting composites. In an attempt to satisfy both of these goals, we present several types of validation results in subsequent sections. Note that the cross-validation approach used here is similar to that used for the Work Suitability Inventory (WSI) analyses reported in Chapter 8.

First, we present validation results based on WPS composites constructed on the full sample (Waves 1 and 2 combined). Basing these composites on the full sample allowed us to obtain the most stable estimates possible for the content and parameters of the weighted composites. After presenting these results, we show estimated validities for models based solely on the Wave 1 sample. We also show cross-validities for WPS composites in Wave 2 by taking the content and weighting underlying Wave 1 WPS composites and applying them to the Wave 2 data. Comparing the Wave 1 validities to the Wave 2 cross-validities allowed us to estimate the amount of shrinkage one might expect to observe from following the modeling processes we used to construct different types of WPS composites (e.g., regression, subjective, unit weighted). It is important to note that comparison of Wave 1 validities and Wave 2 cross-validities will only provide a *rough* estimate of how well the full sample WPS composites would be expected to cross-validate. First, all else being equal, the validity of the full sample WPS composites should be more stable than those based solely on Wave 1 data (due to a larger sample size). Also, given that the full sample and Wave 1 sample only partially overlap, the content and weighing of the full sample and Wave 1 WPS composites may not be identical (even for those composites targeting the same criterion).

Results

Table 10.1 shows descriptive statistics and internal consistency reliability estimates for the WPS scale and facet scores in the full sample. With the potential exception of the Clear Procedures facet of Conventional interests ($\alpha = .63$), and the Prestige facet of Enterprising interests ($\alpha = .68$), all WPS scales exhibited adequate levels of internal consistency (i.e., $\alpha > .70$) and variability.

Table 10.1. Descriptive Statistics for WPS Scales and Facets

Scale/Facet	*k*	α	*M*	*SD*
Realistic Interests Scale	13	0.90	3.28	0.82
Mechanical Facet	5	0.91	3.13	1.07
Physical Facet	6	0.88	3.41	0.92
Investigative Interests Scale	12	0.86	3.23	0.68
Conduct Research Facet	6	0.79	2.82	0.83
Critical Thinking Facet	6	0.85	3.65	0.77
Artistic Interests Scale	12	0.86	2.85	0.74
Artistic Activities Facet	8	0.85	2.46	0.87
Creativity Facet	4	0.84	3.64	0.84
Social Interests Scale	10	0.85	3.46	0.71
Help Others Facet	4	0.72	3.29	0.84
Work with Others Facet	3	0.76	3.56	0.87
Enterprising Interests Scale	13	0.82	3.22	0.61
High Profile Facet	4	0.75	2.52	0.89
Lead Others Facet	3	0.76	3.56	0.85
Prestige Facet	4	0.68	3.71	0.75
Conventional Interests Scale	12	0.81	3.14	0.64
Clear Procedures Facet	3	0.63	3.70	0.80
Detail Orientation Facet	3	0.73	3.70	0.82
Information Management Facet	6	0.82	2.69	0.86

Note. $n = 766$. k = Number of items on scale/facet. α = Cronbach's alpha.

Construct Validity

Table 10.2 shows raw zero-order intercorrelations among the WPS scales and facets. One common way to establish construct validity evidence for the WPS would be to examine it in relation to an established measure of the RIASEC interests. Fortunately, as part of the Select21 concurrent validation effort, the Department of Defense's Career Exploration Program Interest Inventory (CEP-II) was also administered to Soldiers. Like the WPS, the CEP-II was designed to assess Holland's six RIASEC dimensions. However, unlike the WPS, the CEP-II has been established as a valid measure of the RIASEC interests by past research. The CEP-II also differs from the WPS in some other key ways, namely (a) its content is more homogeneous than the WPS, as its items reflect work-related and non-work related activities only (not interest in work environments or learning opportunities); (b) it is based on a 3- point scale of liking (not a 5-point importance scale); and (c) it was developed for vocational counseling (not personnel selection). Despite these differences, comparing the pattern of correlations between these measures could inform construct validity judgments regarding the WPS.

Table 10.2. Intercorrelations among WPS Scales and Facets

Scale/Facet	1	1a	1b	2	2a	2b	3	3a	3b	4	4a	4b	5	5a	5b	5c	6	6a	6b
1. Realistic Interests Scale																			
1a. Mechanical Facet	**.83**																		
1b. Physical Facet	**.82**	**.39**																	
2. Investigative Interests Scale	**.15**	**.14**	**.14**																
2a. Conduct Research Facet	.05	**.09**	.02	**.87**															
2b. Critical Thinking Facet	**.21**	**.15**	**.22**	**.85**	**.48**														
3. Artistic Interests Scale	**.13**	**.18**	.05	**.48**	**.50**	**.32**													
3a. Artistic Activities Facet	**.10**	**.16**	.02	**.35**	**.46**	**.13**	**.94**												
3b. Creativity Facet	**.14**	**.14**	**.09**	**.54**	**.37**	**.56**	**.69**	**.39**											
4. Social Interests Scale	**.15**	.01	**.25**	**.52**	**.40**	**.50**	**.31**	**.21**	**.39**										
4a. Help Others Facet	-.03	**-.10**	.06	**.48**	**.43**	**.39**	**.34**	**.28**	**.31**	**.86**									
4b. Work with Others Facet	**.29**	**.11**	**.36**	**.33**	**.20**	**.37**	**.16**	.05	**.31**	**.81**	**.46**								
5. Enterprising Interests Scale	**.16**	.06	**.24**	**.62**	**.50**	**.56**	**.41**	**.30**	**.45**	**.60**	**.48**	**.48**							
5a. High Profile Facet	.00	.04	-.01	**.47**	**.56**	**.24**	**.47**	**.49**	**.21**	**.30**	**.30**	**.18**	**.72**						
5b. Lead Others Facet	**.26**	.06	**.38**	**.40**	**.22**	**.47**	**.16**	.05	**.32**	**.60**	**.43**	**.55**	**.74**	**.28**					
5c. Prestige Facet	**.13**	.03	**.20**	**.40**	**.23**	**.48**	**.17**	.03	**.39**	**.46**	**.34**	**.38**	**.74**	**.23**	**.50**				
6. Conventional Interests Scale	**.18**	**.17**	**.14**	**.58**	**.51**	**.48**	**.29**	**.25**	**.25**	**.48**	**.42**	**.34**	**.61**	**.50**	**.42**	**.43**			
6a. Clear Procedures Facet	**.24**	**.17**	**.24**	**.42**	**.22**	**.51**	.04	**-.08**	**.27**	**.41**	**.29**	**.36**	**.40**	**.08**	**.38**	**.47**	**.69**		
6b. Detail Orientation Facet	**.30**	**.22**	**.28**	**.49**	**.26**	**.59**	.07	**-.08**	**.34**	**.42**	**.26**	**.41**	**.43**	**.09**	**.43**	**.46**	**.64**	**.86**	
6c. Information Management Facet	.01	**.09**	-.04	**.50**	**.56**	**.28**	**.39**	**.41**	**.17**	**.35**	**.36**	**.18**	**.52**	**.64**	**.27**	**.21**	**.84**	**.27**	**.23**

Note. $n = 766$. Statistically significant correlations are bolded ($p < .05$, two-tailed).

Table 10.3 shows correlations between the WPS and CEP-II. The pattern of correlations shown provides construct validity evidence for the WPS. Specifically, the average mono-trait, hetero-method (scale-level) correlation was .56, whereas the average hetero-trait, mono-method (WPS scale-level) correlation was .38 and the average hetero-trait, hetero-method (scale-level) correlation was .19.[31] Although the average mono-trait, hetero-method correlation was not very large (.56), it is important to remember the differences between the CEP-II and WPS mentioned above. In addition to those general differences, there are also content differences between these measures at the facet level (Van Iddekinge, Putka et al., 2005). For example, whereas the WPS Realistic scale has a facet that assesses physical interests, the CEP-II Realistic scale does not. This is consistent with correlations in Table 10.3 which show the CEP-II Realistic scale correlated more with the WPS Mechanical Facet score (r = .59) than the WPS Physical Facet score (r = .36).

Table 10.3. Correlations between WPS Scores and CEP-II Scale Scores

	CEP-II Scale					
WPS Scale/Facet	R	I	A	S	E	C
Realistic Interests Scale	**.57**	.05	**-.09**	-.05	**-.13**	**-.11**
Mechanical Facet	**.59**	.04	**-.10**	**-.13**	**-.15**	**-.09**
Physical Facet	**.36**	.05	-.04	.06	-.05	**-.08**
Investigative Interests Scale	**.12**	**.55**	**.31**	**.35**	**.40**	**.29**
Conduct Research Facet	**.08**	**.56**	**.30**	**.30**	**.33**	**.29**
Critical Thinking Facet	**.14**	**.38**	**.24**	**.31**	**.36**	**.20**
Artistic Interests Scale	**.10**	**.33**	**.60**	**.23**	**.28**	**.21**
Artistic Activities Facet	**.09**	**.30**	**.60**	**.19**	**.23**	**.20**
Creativity Facet	**.08**	**.26**	**.33**	**.22**	**.28**	**.13**
Social Interests Scale	.02	**.26**	**.24**	**.58**	**.37**	**.21**
Help Others Facet	-.05	**.30**	**.29**	**.59**	**.38**	**.24**
Work with Others Facet	.06	**.10**	**.10**	**.37**	**.21**	.06
Enterprising Interests Scale	.04	**.28**	**.26**	**.34**	**.56**	**.36**
High Profile Facet	.00	**.27**	**.31**	**.21**	**.53**	**.40**
Lead Others Facet	**.10**	**.16**	**.10**	**.35**	**.36**	**.20**
Prestige Facet	.00	**.15**	**.12**	**.22**	**.33**	**.20**
Conventional Interests Scale	**.10**	**.25**	**.16**	**.25**	**.37**	**.54**
Clear Procedures Facet	**.10**	**.13**	.01	**.20**	**.19**	**.22**
Detail Orientation Facet	**.14**	**.15**	.03	**.17**	**.19**	**.16**
Information Management Facet	.04	**.26**	**.22**	**.21**	**.39**	**.62**

Note. n = 514. Mono-trait, hetero-method correlations are boxed. R = CEP-II Realistic Interests Scale. I = CEP-II Investigative Interests Scale. A = CEP-II Artistic Interests Scale. S = CEP-II Social Interests Scale. E = CEP-II Enterprising Interests Scale. C = CEP-II Conventional Interests Scale. Statistically significant correlations are bolded (p < .05, one-tailed).

Criterion-Related Validity Estimates

The previous section provided details on basic psychometric properties of the WPS scales and facets. These scales and facets (along with data from the AES) provide the basis for the WPS composites discussed in this section. Table 10.4 shows criterion-related validity estimates for

[31] The average hetero-trait, mono-method correlation reflects the average of WPS scale intercorrelations from Table 10.2.

WPS composites in the full sample.[32] The table shows both uncorrected and corrected criterion-related validity estimates for each of the 10 Select21 criteria. Analysis details are provided in Chapter 6. Criterion-related validity estimates for the "weighted" composites (i.e., regression, subjective, and unit weighted composites) are not adjusted for shrinkage due to the issues discussed earlier. Later sections of this chapter will present validity estimates by sample to address the issue of how well the weighted composites cross-validated.

The results in Table 10.4 indicate that the WPS has substantial promise as a predictor of the Select21 criteria, particularly the attitudinal criteria. Good levels of validity were also found for predicting the Achievement and Effort performance criterion. The fact that the WPS

Table 10.4. Criterion-Related Validity Estimates for WPS Composites in the Full Sample

	Performance Criteria					Attitudinal Criteria				
Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
D^2 Fit Index	**-.08**	**-.09**	**-.12**	.00	-.04	**-.26**	**-.27**	**.18**	**-.19**	**-.14**
Pearson *r* Fit Index	.04	**.14**	**.11**	.00	.05	**.29**	**.28**	**-.22**	**.19**	**.11**
Regression	**.21**	**.31**	**.25**	.	**.13**	**.40**	**.45**	**.40**	**.33**	**.34**
Subjective	.	.	**.22**	.	.	**.39**	**.43**	**-.37**	**.31**	**.32**
Unit	**.18**	**.30**	**.20**	.	**.13**	**.36**	**.40**	**-.35**	**.31**	**.31**
Unit AE	**.15**	**.30**	**.11**	**.12**	**.14**	**.21**	**.26**	**-.14**	**.18**	**.20**
Subjective AFit	**.10**	**.20**	**.20**	**.07**	**.10**	**.36**	**.43**	**-.30**	**.31**	**.31**
	Corrected Validity Estimates									
D^2 Fit Index	-.03	-.07	-.12	.01	-.02	-.28	-.29	.19	-.21	-.15
Pearson *r* Fit Index	-.03	.11	.11	-.03	.01	.31	.31	-.23	.20	.12
Regression	.33	.34	.26	.	.22	.43	.50	.46	.35	.36
Subjective	.	.	.23	.	.	.41	.47	-.43	.33	.34
Unit	.29	.34	.21	.	.22	.39	.44	-.41	.33	.33
Unit AE	.18	.34	.11	.21	.20	.22	.29	-.18	.18	.21
Subjective AFit	.03	.17	.20	.08	.08	.38	.47	-.31	.33	.33

Note. n = 546 (AE criterion), n = 731-732 (all other performance criteria), n = 703-723 (attitudinal criteria). Regression = Regression weighted composite score specific to each criterion optimized in the full sample. Subjective = Subjectively weighted composite score specific to each criterion based on regression analyses in the full sample. Unit = Unit weighted composite score specific to each criterion based on regression analyses in the full sample. Unit AE = Unit weighted composite score formed based on the AE performance criterion. Subjective AFit = Subjectively weighted composite score formed based on the Perceived Fit with the Army (AFit) attitudinal criterion. Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

[32] As Table 10.4 shows, subjectively weighted composites were not constructed for some performance criteria due to their lack of differentiation from unit weighted composites. Thus, validity estimates for subjectively weighted composites are missing for several criteria in Table 10.4. Furthermore, criterion related-related validity estimates for regression, subjectively weighted, and unit weighted composites are not provided for Teamwork because a decision was made not to "model" this criterion due to its unreliability (cf. Chapter 3).

predicted Achievement and Effort performance (which is more of a will-do performance criterion) is consistent with recent research suggesting a link between interests and job performance (Hogan & Hogan, 1996; Quintela, 2003).

With regard to the magnitude of the criterion-related validity estimates for predicting attitudinal criteria, they were fairly impressive in both an absolute sense and compared to past literature. For example, in Project A, the average unadjusted multiple correlation between the six composites from the Army Vocational Interest Career Examination (AVOICE) and Satisfaction with the Army (across MOS) was .14 based on a *longitudinal* validation sample (Knapp & Carter, 2003).[33] In contrast, the regression weighted WPS composite targeting Satisfaction with the Army had an uncorrected validity of .40. Although the WPS appeared to show much more validity then the AVOICE, caution should be taken not to overinterpret this difference in validity given the concurrent nature of the Select21 sample.

While comparing favorably to past Army research, these results also compare favorably to past research in the civilian vocational interest and P-E fit literatures. For example, meta-analyses have estimated the relationship between vocational interest congruence indexes and satisfaction to be roughly .20 (Assouline & Meir, 1987; Tranberg, Slane, & Ekeberg, 1993). Furthermore, Kristof-Brown et al. (2005) reported meta-analytic estimates of .29 and -.19, respectively, for the criterion-related validity of indirect, objective measures of P-E fit for satisfaction and intentions to quit (similar to attrition cognitions). The finding of larger validity estimates for the "weighted" Select21 composites is not surprising given that the meta-analytic estimates were primarily based on relations between similarity indexes and criteria. These findings provide further evidence that profile similarity indexes such as D^2 and Pearson r commonly used in the P-E fit literature artificially constrain observed P-E-C relations.[34]

Despite the merits of regression weighted composites discussed in the introduction, results in Table 10.4 show that simple, subjectively weighted and unit weighted composites exhibit comparable levels of validity to their regression weighted counterparts. Upon cross-validation we would expect these validities to become even more similar.

Given the similarity between the attitudinal criteria, perhaps it is not surprising that we were also able to obtain good levels of validity by using composites optimized on one criterion as predictors of other criteria. For example, the subjectively weighted composite targeting Perceived Fit with the Army had criterion-related validities for predicting all other attitudinal criteria that exceeded .30 in magnitude. In light of such results, and in the interest of creating a more parsimonious set of WPS predictors, we limited our attention to only two of the 26 composites summarized in Table 10.4 for subsequent cross-instrument analyses in this report (see Chapters 13-15), namely the Unit Achievement and Effort and Subjective Perceived Army Fit composites. Of any of the WPS composites, these two had the highest absolute validity (on average) for predicting the performance and attitudinal criteria, respectively.

---

[33] The AVOICE was a RIASEC-based vocational interest measure developed in Project A.

[34] We acknowledge that unlike the fit index-based composites, the other WPS composites were at least partially optimized on the criteria. As such, upon cross-validation we would expect to see less of a difference between the validity of the WPS composites based on fit indexes and those optimized on the criteria. Analyses presented later in this chapter provide a rough indication of how much smaller these differences may become upon cross-validation.

Composition of WPS Composites

One of the most interesting aspects of the present findings is the composition of the weighted WPS composites. In earlier sections of this chapter, we noted that the modeling that underlies the regression weighted composites would take place at the scale-level so that we could incorporate WPS and AES data using methods described in Putka (2005). As it turns out, we were able to achieve far better prediction of the criteria by ignoring the AES data altogether, and modeling the criteria as a function of the WPS facet-level scores. Thus, as Table 10.5 reveals, the weighted WPS composites consist solely of WPS facet-level scores. This finding casts serious doubt on a fundamental assumption underlying the construction and validation of interest-based P-E fit measures, namely that it is necessary to incorporate environment-side information into the prediction composite (be it a fit index, or a some regression-based composite) to obtain good prediction (Ployhart et al., 2006). Indeed, comparing the validity of the fit indexes in Table 10.4 to the validity of the weighted composites suggests that we consistently obtained higher

Table 10.5. Composition of Weighted WPS Composites

	Performance Criteria				Attitudinal Criteria				
Scale/Facet	GTP	AE	PF	FXP	ASat	AFit	ACog	CInt	FAA
Realistic Interests									
Mechanical Facet	.	.	-0.12	.	.	.	.	.	.
Physical Facet	.	.	0.26[b]	.	0.31[b]	0.28[b]	-0.28[b]	0.18[a]	0.22[a]
Investigative Interests									
Conduct Research Facet	.	.	0.07[a]	.	.	.	.	.	0.08[c]
Critical Thinking Facet	0.21[a]	0.10[a]	.	0.13[a]	.	.	-0.14[a]	.	0.11[a]
Artistic Interests									
Artistic Activities Facet	.	-0.18[a]	.	.	-0.11[a]	.	.	.	.
Creativity Facet	.	.	.	.	.	-0.10	0.13	-0.07	.
Social Interests									
Help Others Facet	.	0.16[a]	.	.	.	0.09[a]	.	.	.
Work with Others Facet	.	.	.	.	0.14[a]	0.12[a]	-0.18[a]	.	.
Enterprising Interests									
High Profile Facet	.	.	.	.	-0.10	-0.09	0.11[a]	-0.08	-0.08
Lead Others Facet	.	.	.	.	.	0.11[a]	.	0.19[a]	0.10[a]
Prestige Facet	.	.	.	.	.	.	.	.	.
Conventional Interests									
Clear Procedures Facet	.	.	.	.	.	0.08[a]	0.09	0.09[c]	.
Detail Orientation Facet	.	0.10[a]	.	.	.	.	.	.	.
Information Management Facet	-0.11	.	.	.	0.13	.	.	.	.

Note. Cell values reflect standardized beta weights for the WPS regression-based composite targeting the given criterion. If no cell value is listed for a given WPS scale/facet, then it means that the WPS scale/facet was not included in the composite for the given criterion. All scales that have superscripts on their standardized beta weights were included in unit weighted and subjectively weighted composites targeting the given criterion. GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

[a] The scale was included in unit weighted and subjectively weighted composites targeting the given criterion and received a weight of +1 or -1 (depending on the direction of its zero-order correlation with the criterion).

[b] The scale was given a weight of 2 in the subjectively weighted composite targeting the given criterion.

[c] The scale was given a weight of 0.5 in the subjectively weighted composite targeting the given criterion.

validities ignoring environment-side data. These findings also suggest that similar to criticisms made with respect to the Big Five personality factors, better prediction may be achieved by using facets of the RIASEC dimensions to predict criteria, rather than using the dimensions themselves (Schneider, Hough, & Dunnette, 1996)

Table 10.5 suggests little consistency in the composition of composites designed to predict the Select21 performance criteria (with the exception of WPS Critical Thinking). Such results provide partial evidence for the discriminant validity of the performance dimensions. For example, the WPS Physical facet had the highest regression weight among facets in the composite targeting Physical Fitness performance. The WPS Critical Thinking facet had the highest regression weight among facets in the composite targeting General Technical Proficiency (i.e., the performance composite with the strongest links to cognitive ability).

On the attitudinal side, there was more consistency in the composition of the composites. For example, the WPS Physical facet played a key role in all of the composites. Such a finding is consistent with past research which has suggested physical fitness (in this case physical interests), plays a key role in understanding the attitudes and behaviors of Soldiers (Strickland, 2005). Several other facets were also included in composites for three or more of the attitudinal criteria. Specifically, the WPS High Profile facet was in the regression weighted composite of all five attitudinal criteria, and the WPS Clear Procedures, Works with Others, Lead Others, and Creativity facets were in regression weighted composites for three of the five criteria. The fact that these characteristics consistently emerged across criteria (both in magnitude and direction) appears consistent with the extent to which those interests that are supported by the Army environment. For example, the Army generally offers Soldiers opportunities to engage in physical activity, clear procedures for accomplishing tasks, and opportunities to work with and lead others, but arguably offers fewer opportunities for creativity and high profile work (at least for first-term Soldiers).

Relations among Composites

The criterion-related validity estimates of many WPS composites were presented in Table 10.3. Table 10.6 shows the correlation between the final two composites we chose to move forward with and the other WPS composites. Not surprisingly, the two final composites were highly related to the other weighted composites that targeted the same criterion (e.g., the unit weighted composite targeting Achievement and Effort correlated .99 with the regression weighted composite targeting Achievement and Effort). In general, both of the final composites were moderately to strongly related to the other composites, with many correlations exceeding .60. This pattern was particularly true for relations between the Subjective Perceived Fit with the Army composite and composites targeting the other attitudinal criteria. This finding is not surprising given the moderate to high correlations observed between the attitudinal criteria in Chapter 3. Interestingly, neither of the final composites was strongly correlated with the Pearson *r* fit index, which suggests that these weighted composites assess something different than similarity between Soldiers' profiles on the WPS and the Army profile based on the AES.

Table 10.6. Correlations between Final WPS Composites and Other WPS Composites

All WPS Composites	Final WPS Composites: Unit AE	Final WPS Composites: Subjective AFit
1. D^2 Fit Index	-.33	-.56
2. Pearson *r* Fit Index	.30	.39
3. Regression General Technical Proficiency	.66	.40
4. Unit General Technical Proficiency	.75	.52
5. Regression Achievement and Effort	.99	.61
6. Unit Achievement and Effort	1.00	.65
7. Regression Physical Fitness	.29	.75
8. Subjective Physical Fitness	.31	.79
9. Unit Physical Fitness	.34	.73
10. Regression Future Expected Performance	.75	.52
11. Unit Future Expected Performance	.75	.52
12. Regression Satisfaction with the Army	.47	.84
13. Subjective Satisfaction with the Army	.45	.81
14. Unit Satisfaction with the Army	.53	.76
15. Regression Perceived Army Fit	.55	.94
16. Subjective Perceived Army Fit	.65	1.00
17. Unit Perceived Army Fit	.72	.98
18. Regression Attrition Cognitions	-.36	-.75
19. Subjective Attrition Cognitions	.51	.84
20. Unit Attrition Cognitions	.60	.81
21. Regression Career Intentions	.52	.86
22. Subjective Career Intentions	.56	.94
23. Unit Career Intentions	.64	.94
24. Regression Future Army Affect	.58	.88
25. Subjective Future Army Affect	.64	.88
26. Unit Future Army Affect	.63	.84

Note. n = 765-766. Correlations that appear in boxes are for those WPS composites that target the same criterion as the WPS composite shown at the top of the given column. All correlations are statistically significant ($p < .05$, one-tailed).

Cross-Validation of Composites

Table 10.7 shows criterion-related validity estimates for WPS composites in the Wave 1 and Wave 2 samples.[35] Unlike Table 10.4, the weighted WPS composites in this table were constructed based on the Wave 1 data only. Thus, the Wave 2 validity estimates represent cross-validities (i.e., criterion-related validities based on applying Wave 1 parameters to Wave 2 data). Interestingly, the weighted WPS composites constructed in Wave 1 retained their validity very well in Wave 2. In fact, for the regression weighted composites targeting Perceived Fit with the Army and Attrition Cognitions, the Wave 2 validities were actually slightly higher than the Wave 1 validities. Furthermore, all of the subjectively weighted and unit weighted composites targeting attitudinal criteria had slightly higher estimated validities in Wave 2 compared to Wave 1. While somewhat surprising, such findings may be understandable given the similarity between the Wave 1 and 2 samples (see Chapter 2).

[35] Like Table 10.4, several values are missing from this table. See Footnote 11 for an explanation of the missing values.

Table 10.7. Criterion-Related Validity Estimates for WPS Composites in the Wave 1 and Wave 2 Samples

	Performance Criteria					Attitudinal Criteria				
Sample/Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
Wave 1 Sample										
D^2 Fit Index	-.07	-.08	**-.15**	.01	-.06	**-.23**	**-.23**	**.17**	**-.16**	**-.15**
Pearson *r* Fit Index	**.07**	**.17**	**.13**	.02	**.09**	**.29**	**.29**	**-.22**	**.20**	**.13**
Regression (W1)	**.19**	**.30**	**.26**	**.**	**.20**	**.42**	**.43**	**.38**	**.32**	**.33**
Subjective (W1)	**.**	**.**	**.25**	**.**	**.**	**.37**	**.40**	**-.33**	**.30**	**.31**
Unit (W1)	**.17**	**.30**	**.23**	**.**	**.10**	**.33**	**.37**	**-.31**	**.30**	**.30**
Wave 2 Sample										
D^2 Fit Index	-.12	**-.14**	-.01	-.04	-.01	**-.34**	**-.37**	**.21**	**-.28**	**-.14**
Pearson *r* Fit Index	-.02	.10	.06	-.03	-.03	**.32**	**.28**	**-.24**	**.19**	.07
Regression (W1)	**.17**	**.26**	**.23**	.	.10	**.36**	**.50**	**.45**	**.29**	**.29**
Subjective (W1)	.	.	**.17**	.	.	**.39**	**.50**	**-.38**	**.36**	**.32**
Unit (W1)	.09	**.27**	**.13**	.	**.17**	**.38**	**.46**	**-.34**	**.33**	**.28**
	Corrected Validity Estimates									
Wave 1 Sample										
D^2 Fit Index	-.02	-.05	-.15	.03	-.04	-.25	-.25	.17	-.17	-.16
Pearson *r* Fit Index	.00	.14	.13	.00	.06	.30	.31	-.22	.21	.14
Regression (W1)	.25	.30	.27	.	.29	.44	.46	.44	.31	.33
Subjective (W1)	.	.	.25	.	.	.38	.43	-.38	.31	.33
Unit (W1)	.20	.30	.23	.	.13	.35	.40	-.36	.31	.33
Wave 2 Sample										
D^2 Fit Index	-.14	-.14	-.02	-.06	.01	-.36	-.41	.24	-.29	-.15
Pearson *r* Fit Index	-.07	.08	.07	-.06	-.10	.34	.32	-.27	.21	.09
Regression (W1)	.33	.30	.24	.	.26	.39	.56	.53	.31	.30
Subjective (W1)	.	.	.18	.	.	.43	.56	-.44	.40	.34
Unit (W1)	.14	.31	.14	.	.28	.41	.52	-.40	.35	.30

Note. n_{Wave1} = 397 (AE criterion), n_{Wave1} = 562-563 (all other performance criteria), n_{Wave1} = 531-550 (attitudinal criteria). n_{Wave2} = 149 (AE criterion), n_{Wave2} = 169 (all other performance criteria), n_{Wave2} = 172-173 (attitudinal criteria). Regression (W1) = Regression weighted composite score specific to each criterion optimized in the Wave 1 sample. Subjective (W1) = Subjectively weighted composite score specific to each criterion based on regression analyses in Wave 1 sample. Unit (W1) = Unit weighted composite score specific to each criterion based on regression analyses in Wave 1 sample. Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Although similarity of Wave 1 and 2 samples may be one explanation for these results, other factors may also account for the findings. One of these factors is sampling error. The Wave 2 sample consists of fewer than 200 Soldiers; as such, these results may simply reflect the particular sample we obtained (in this case, we may be on the fortunate side of sampling error). Another possibility stems from differential amounts of range restriction within the samples. Specifically, we observed that there was slightly more variation (on average) in Soldiers' WPS composite scores in Wave 2 than in Wave 1. All else being equal, higher variances on the WPS

composites in Wave 2 would equate to higher estimated validities (or in this case, less shrinkage). Another explanation might be the modeling process itself. Although regression analyses were used to create the regression weighted composites, careful attention was paid to the theoretical meaningfulness of relationships uncovered by this modeling process. In general, we were very conservative when it came to including terms whose weights were difficult to reconcile with theory or that appeared to be driven by a few influential cases. In such cases, we tended to use a model that was more consistent with theory at the expense of potentially sacrificing a few hundredths of a point on a validity coefficient. Thus, consistent with the suggestions made by Putka (2005), the modeling process was not purely empirical, and as such, may be less subject to shrinkage than a process driven entirely by the data.

Incremental Validity Estimates

In the previous section, we provided evidence for the criterion-related validity of the WPS. Here we focus on the degree to which it increments the validity of the AFQT. Table 10.8 shows incremental validity estimates for the WPS composites in the full sample.[36] These estimates show that the WPS has a substantial level of incremental validity over the AFQT for predicting the attitudinal criteria. This finding is not surprising given the general lack of validity of the AFQT for predicting attitudinal criteria and the good validity of the WPS for predicting attitudes. With regard to the performance criteria, the incremental validity of the WPS composites over the AFQT was notable for the Achievement and Effort and Physical Fitness performance composites, but not for the General Technical Proficiency composite. This finding is consistent with our expectations, and indeed, the composition of the weighted WPS composites themselves. As alluded to earlier, we believed the strongest predictor of General Technical Proficiency would be AFQT scores because General Technical Proficiency reflects more of a "can-do" performance criterion. As such, predictors that assess motivation-related determinants of performance (such as the WPS) may have little to offer over the AFQT for predicting General Technical Proficiency. Nevertheless, when it comes to more "will-do" performance criteria such as Achievement and Effort, we would expect the WPS to increment the AFQT, and indeed it does. The significant increment observed for predicting Physical Fitness makes sense for two reasons. First, we would expect that measures of cognitive ability such as the AFQT to have little to do with physical fitness performance (and indeed in this sample the correlation was zero); thus, the potential to observe incremental validity is present. Second, given that that the WPS composites targeting Physical Fitness includes the WPS Physical facet score as a key element, it is not surprising that those composites, along with others which include the WPS Physical facet (e.g., Subjective Perceived Fit with the Army), incremented the AFQT for predicting Physical Fitness performance.

Subgroup Differences

Tables 10.9 and 10.10 show mean final WPS composite scores by gender and race/ethnicity, respectively. Though two statistically significant differences were found (in both cases the minority groups were higher), the magnitudes of these effects sizes were relatively small, as no effect sizes exceeded 0.31 in magnitude.

[36] Like Tables 10.4 and 10.7, several values are missing from this table. See Footnote 11 for an explanation of the missing values.

Table 10.8. Incremental Validity Estimates for WPS Composites in the Full Sample

	Performance Criteria					Attitudinal Criteria				
Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Incremental Validity Estimates									
AFQT	**.31**	**.17**	.00	**.07**	**.18**	-.01	.01	**-.13**	**-.07**	-.04
D^2 Fit Index	**.02**	**.03**	**.11**	.00	.01	**.25**	**.26**	**.10**	**.13**	**.11**
Pearson *r* Fit Index	**.01**	**.07**	**.11**	.00	.01	**.28**	**.28**	**.15**	**.12**	**.08**
Regression	**.03**	**.17**	**.25**	.	**.02**	**.39**	**.45**	**.30**	**.27**	**.30**
Subjective	.	.	**.22**	.	.	**.37**	**.43**	**.27**	**.25**	**.29**
Unit	**.03**	**.17**	**.20**	.	**.02**	**.35**	**.40**	**.25**	**.24**	**.27**
Unit AE	**.03**	**.17**	**.10**	**.07**	**.04**	**.20**	**.25**	**.06**	**.13**	**.17**
Subjective AFit	**.04**	**.12**	**.20**	**.03**	**.04**	**.35**	**.43**	**.22**	**.24**	**.28**
	Corrected Incremental Validity Estimates									
AFQT	.53	.30	.00	.19	.37	-.02	.01	-.24	-.11	-.06
D^2 Fit Index	.01	.01	.09	.00	.00	.25	.28	.08	.10	.08
Pearson *r* Fit Index	.00	.04	.08	.00	.01	.28	.30	.13	.09	.04
Regression	.02	.14	.25	.	.01	.40	.50	.30	.24	.30
Subjective	.	.	.22	.	.	.39	.47	.27	.22	.28
Unit	.01	.13	.20	.	.01	.36	.43	.24	.21	.26
Unit AE	.02	.13	.08	.08	.03	.19	.27	.04	.10	.15
Subjective AFit	.02	.09	.19	.03	.03	.36	.47	.21	.22	.27

Note. $n = 524$ (AE criterion), $n = 707$ (all other performance criteria), $n = 677$-699 (attitudinal criteria). Cell values for the AFQT represent zero-order correlations between the AFQT and the given criterion (shown for reference). Uncorrected incremental estimates reflect the difference between the multiple *R* obtained when regressing the criterion on both the given composite and AFQT versus the *R* obtained when regressing the criterion only on the AFQT. Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Statistically significant incremental validities are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Table 10.9. Final WPS Composite Scores by Gender

		Male		Female	
WPS Composite	d_{FM}	*M*	*SD*	*M*	*SD*
Unit AE	**0.26**	2.03	0.49	2.16	0.44
Subjective AFit	-0.14	4.20	0.72	4.09	0.70

Note. $n_{Male} = 683$. $n_{Female} = 82$. d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females – mean of males)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 10.10. Final WPS Composite Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
WPS Composite	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unit AE	0.13	0.13	2.03	0.47	2.09	0.50	2.01	0.48	2.07	0.52
Subjective AFit	-0.06	**0.31**	4.19	0.71	4.14	0.72	4.13	0.70	4.35	0.76

Note. n_{White} = 546. n_{Black} = 146. $n_{White\ Non\text{-}Hispanic}$ = 430. $n_{Hispanic}$ = 145. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites)/*SD* of Whites. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Differential Prediction

Tables 10.11 through 10.13 present the results of differential prediction analyses for the final WPS composites. Table 10.11 shows results for gender, Table 10.12 for race, and Table 10.13 for race/ethnicity.[37] Overall, the results indicate some evidence of intercept bias and differential prediction (i.e., slope bias) depending on the criterion, WPS composite, and demographic variable considered. In light of these findings, we discuss results from each of the tables in turn, and focus only on interpreting results for the criteria each WPS composite was meant to predict (Unit Achievement and Effort [Unit AE]—performance criteria; Subjective Fit with the Army [Subjective AFit]—attitudinal criteria).

Table 10.11. Differential Prediction Results by Gender for Final WPS Composites

	Unit AE WPS Composite					Subjective AFit WPS Composite				
	Gender	WPS *b*		*r* by Gender		Gender	WPS *b*		*r* by Gender	
Criterion	*b*	M	F	M	F	*b*	M	F	M	F
Performance Criteria										
General Technical Proficiency	-0.04	0.07	0.16	**.13**	**.29**	0.01	0.05	0.10	**.09**	.19
Achievement and Effort	**0.19**	0.15	0.16	**.29**	**.28**	**0.25**	0.11	0.06	**.23**	.11
Physical Fitness	-0.18	0.07	0.21	**.10**	**.23**	-0.07	**0.13**	**0.34**	**.17**	**.39**
Teamwork	**0.18**	0.06	0.14	**.11**	.21	**0.24**	0.03	0.13	.06	.19
Future Expected Performance	**0.17**	0.08	0.10	**.13**	.15	0.20	0.07	0.05	**.11**	.08
Attitudinal Criteria										
Satisfaction with the Army	**-0.24**	0.17	0.18	**.22**	**.23**	-0.15	0.27	0.33	**.34**	**.45**
Perceived Army Fit	-0.11	0.21	0.30	**.26**	**.33**	0.03	0.35	0.46	**.41**	**.53**
Attrition Cognitions	**0.39**	-0.14	-0.20	**-.15**	-.18	**0.29**	-0.28	-0.33	**-.29**	**-.32**
Career Intentions	-0.05	0.20	0.25	**.18**	.20	0.08	0.34	0.47	**.30**	**.40**
Future Army Affect	**-0.33**	0.20	0.20	**.21**	.21	**-0.24**	0.29	0.30	**.30**	**.33**

Note. $n_{Regression}$ = 545-731. n_{Male} = 481-657. n_{Female} = 64-80. Gender *b* = Unstandardized regression weight for gender (0 = male, 1 = female). WPS *b* = Unstandardized regression weight for the given WPS composite for males and females. *r* by Gender = Correlation between the given WPS composite and the given criterion for each gender. Regression weights for males and females are bolded if the WPS-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

[37] All WPS composite scores were standardized prior to conducting these analyses to ease interpretation of the unstandardized regression weights presented in these tables.

Table 10.11 reveals little evidence of slope bias for the Unit AE composite and Subjective AFit composites by gender. On the other hand, intercept bias was apparent when using Unit AE to predict Achievement and Effort, Teamwork, and Future Expected Performance, and when using Subjective AFit to predict Attrition Cognitions and Future Army Affect. In the case of the Unit AE composite, women had Achievement and Effort, Teamwork, and Future Expected Performance scores that were roughly 0.17 to 0.19 points higher than men (at mean levels of the Unit AE composite). These findings suggest that using the Unit AE composite scores would tend to underpredict females' performance on Achievement and Effort, Teamwork, and Future Expected Performance if a common prediction equation were used for all respondents. In the case of the Subjective AFit composite, women had Attrition Cognitions scores that were roughly 0.29 points higher than men and Future Army Affect scores that were roughly 0.25 points lower than men (at mean levels of the Unit AE composite). These findings suggest that using Subjective PFit composite scores would tend to underpredict females' Attrition Cognitions and overpredict their Future Army Affect if a common prediction equation was used.

Table 10.12. Differential Prediction Results by Race for Final WPS Composites

	Unit AE WPS Composite					Subjective AFit WPS Composite				
	Race	WPS *b*		*r* by Race		Race	WPS *b*		*r* by Race	
Criterion	*b*	W	B	W	B	*b*	W	B	W	B
Performance Criteria										
General Technical Proficiency	**-0.27**	0.10	0.09	**.18**	**.21**	**-0.25**	0.04	0.12	.06	**.26**
Achievement and Effort	**-0.19**	0.18	0.12	**.35**	**.21**	**-0.18**	0.12	0.05	**.25**	.09
Physical Fitness	-0.01	0.10	0.06	**.12**	.08	0.02	0.15	0.19	**.19**	**.25**
Teamwork	-0.04	0.08	0.07	**.14**	.11	-0.03	0.03	0.05	.05	.08
Future Expected Performance	-0.18	0.12	0.07	**.18**	.13	-0.16	0.06	0.11	**.09**	**.19**
Attitudinal Criteria										
Satisfaction with the Army	-0.08	0.21	0.06	**.26**	.08	-0.06	0.31	0.29	**.39**	**.38**
Perceived Army Fit	-0.14	**0.28**	**0.06**	**.33**	.07	-0.13	0.41	0.27	**.48**	**.34**
Attrition Cognitions	**0.39**	-0.18	-0.11	**-.18**	-.12	**0.36**	-0.33	-0.29	**-.32**	**-.30**
Career Intentions	0.10	**0.27**	**-0.02**	**.24**	-.01	0.11	**0.45**	**0.11**	**.39**	.10
Future Army Affect	**-0.18**	**0.25**	**0.05**	**.27**	.05	-0.16	**0.34**	**0.15**	**.36**	.15

Note. $n_{\text{Regression}}$ = 496-661. n_{White} = 395-525. n_{Black} = 101-136. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black). WPS *b* = Unstandardized regression weight for the given WPS composite for Whites and Blacks. *r* by Race = Correlation between the given WPS composite and the given criterion for each race. Regression weights for Whites and Blacks are bolded if the WPS-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 10.12 reveals little evidence of slope bias for the Unit AE composite by race. Nevertheless, slope bias was apparent for the Subjective AFit composite when using Career Intentions and Future Army Affect. Specifically, the Subjective AFit score was more predictive of Career Intentions and Future Army Affect for White Soldiers (Career Intentions: $b = .45$, $r = .39$; Future Army Affect: $b = .34$, $r = .36$) than for Black Soldiers (Career Intentions: $b = .11$, $r = .10$; Future Army Affect: $b = .15$, $r = .15$). Intercept bias was apparent when using Unit AE to predict General Technical Proficiency and Achievement and Effort, and when using Subjective AFit to predict Attrition Cognitions. In the case of Unit AE, Black Soldiers had General

Technical Proficiency and Achievement and Effort scores that were roughly 0.27 and 0.19 points (respectively) lower than White Soldiers (at mean levels of the Unit AE composite). These findings suggest that using Unit AE scores would tend to overpredict Black Soldiers' performance on General Technical Proficiency and Achievement and Effort if a common prediction equation were used. In the case of the Subjective AFit composite, Black Soldiers had Attrition Cognitions scores that were roughly 0.36 points higher than White Soldiers (at mean levels of the Subjective AFit composite). These findings suggest that using Subjective AFit WPS scores would tend to underpredict Black Soldiers' Attrition Cognitions if a common prediction equation were used.

Table 10.13. Differential Prediction Results by Ethnic Group for Final WPS Composites

	Unit AE WPS Composite					Subjective AFit WPS Composite				
		WPS *b*		*r* by Ethnicity			WPS *b*		*r* by Ethnicity	
Criterion	Eth *b*	W	H	W	H	Eth *b*	W	H	W	H
Performance Criteria										
General Technical Proficiency	-0.06	0.11	0.04	**.19**	.09	-0.07	0.03	0.06	.06	.13
Achievement and Effort	0.03	0.18	0.18	**.36**	**.36**	0.04	0.13	0.07	**.26**	.14
Physical Fitness	0.09	0.10	0.09	**.12**	.13	0.05	0.13	0.18	**.17**	**.25**
Teamwork	**0.15**	0.07	0.07	**.12**	.13	**0.14**	0.02	0.05	.03	.11
Future Expected Performance	0.06	0.12	0.07	**.17**	.13	0.06	0.05	0.06	.07	.11
Attitudinal Criteria										
Satisfaction with the Army	0.10	**0.24**	**0.05**	**.31**	.06	0.05	**0.35**	**0.15**	**.43**	**.19**
Perceived Army Fit	0.07	0.30	0.20	**.35**	**.26**	0.01	**0.44**	**0.25**	**.51**	**.32**
Attrition Cognitions	0.02	-0.19	-0.16	**-.18**	**-.17**	0.08	-0.35	-0.21	**-.34**	**-.22**
Career Intentions	0.00	0.29	0.18	**.25**	**.17**	-0.09	0.48	0.31	**.39**	**.29**
Future Army Affect	0.17	0.25	0.16	**.27**	**.18**	0.12	0.35	0.21	**.37**	**.25**

Note. $n_{\text{Regression}}$ = 413-552. $n_{\text{White non-Hispanic}}$ = 312-412. $n_{Hispanic}$ = 101-140. Eth *b* = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). WPS *b* = Unstandardized regression weight for the given WPS composite for White non-Hispanics and Hispanics. *r* by Ethnicity = Correlation between the given WPS composite and the given criterion for each race. Regression weights for White non-Hispanics and Hispanics are bolded if the WPS-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 10.13 reveals no evidence of intercept bias for the Subjective AFit composite by race/ethnicity, and some evidence of intercept bias for the Unit AE composite when predicting Teamwork performance (Hispanics were slightly higher on Teamwork than were white non-Hispanics). Although no evidence of slope bias was apparent for the Unit AE composite, slope bias was apparent for the Subjective AFit composite when using it to predict Satisfaction with the Army and Perceived Army Fit. Specifically, the Subjective AFit composite score was more predictive of Satisfaction with the Army and Perceived Army Fit for White non-Hispanic Soldiers (Satisfaction with the Army: $b = .35$, $r = .43$; Perceived Army Fit: $b = .44$, $r = .51$) than for Hispanic Soldiers (Satisfaction with the Army: $b = .15$, $r = .19$; Perceived Army Fit: $b = .25$, $r = .32$).

Discussion

Based on the results presented in this chapter, the WPS appears to be a reliable and construct-valid measure of the RIASEC interest dimensions. Furthermore, the final WPS composites we recommend considering for future use in Soldier selection exhibit minimal mean group differences across genders and racial/ethnic groups. Examination of the criterion-related validity of the WPS suggests it has substantial promise for predicting various attitudinal criteria found to be key precursors of attrition and re-enlistment behavior (Strickland, 2005). Results also indicate that the WPS has promise for predicting Achievement and Effort and Physical Fitness performance above and beyond the AFQT. The findings with regard to the criterion-related validity of the WPS are generally stronger than those found in past Army research with other interest measures, as well as civilian research on vocational interests and P-E fit measures. As noted previously, part of the reason for the success of the WPS may be the more rigorous approach taken to modeling person-environment-criterion relations than is typically seen in the research literature.

While the aforementioned results are promising, there are some causes for concern with the WPS. Specifically, analyses revealed some evidence that predictive biases may result from using the WPS in selection. In some cases, biases such as the intercept differences found across genders are due primarily to the subgroup differences on the criteria of interest rather than to the WPS itself (see Chapters 3 through 5). In other cases, the observed biases may be more problematic. For example, we found that the Subjective AFit WPS composite was more predictive of career intentions and future Army affect for White Soldiers compared to Black Soldiers, and more predictive of satisfaction with the Army in general and perceived fit with the Army for White non-Hispanic Soldiers compared to Hispanic Soldiers.

With regard to the future use of the WPS, we suggest several steps be taken. First, we suggest that the WPS be administered experimentally in an operational selection context and a longitudinal validation effort be conducted. Although this chapter has clearly demonstrated the WPS has validity for predicting criteria in a concurrent sample, there are simply too many factors at play in an operational context (e.g., response distortion) which may attenuate the validity observed here to draw strong conclusions regarding how well the WPS would perform operationally. Indeed previous Army research has demonstrated that the magnitude of differences between the psychometric properties of non-cognitive measures administered in operational and concurrent contexts can be substantial (Knapp, Waters, & Heggestad, 2002).

Another consideration for future use of the WPS should be its potential utility for classification. In developing interest-based P-E fit measures for Select21, our primary focus was on assessing person-Army fit with regard to work-related interests. This method runs contrary to how vocational interest measures have traditionally been used in the vocational counseling and P-E fit literatures. Typically, interest measures have been used to assess fit to a particular occupation, vocation, or job (e.g., an MOS). We deviated from this tradition due to a generally held belief that the Army work environment provides a strong context that permeates the jobs of all first-term Soldiers, regardless of MOS. The fact that the WPS was quite predictive of Army-wide criterion measures examined in this chapter (irrespective of MOS) suggests that this approach was merited. Nevertheless, these results should not be interpreted as meaning that measures of interest-related MOS fit would fail to increment the validity of the interests-related

Army fit composites when predicting MOS-specific criteria (e.g., satisfaction with MOS, perceived fit with MOS, MOS-specific performance). As such, we suggest future Army research, such as the research being conducted as part of ARI's Army Class project, assess whether WPS composites optimized within MOS offer any increment in validity over the more general person-Army fit composites described in this chapter.

CHAPTER 11: WORK VALUES INVENTORY

Dan J. Putka
HumRRO

Overview

Several P-E fit predictor measures were developed in Select21 to predict the attitudinal precursors of attrition and re-enlistment, two criteria of particular interest to the Army (Van Iddekinge, Putka, & Sager, 2005). In the previous chapter, we described the validation of an interests-based P-E fit predictor measure based on Holland's (1985) RIASEC taxonomy of vocational interests. In this chapter, we describe validation results for the Work Values Inventory, a work values-based P-E fit predictor measure derived from Dawis and Lofquist's (1984) Theory of Work Adjustment.

Instrument Description

The Work Values Inventory (WVI) is a computerized card sorting task in which respondents order a list of 28 occupational reinforcers in terms of importance to them on their ideal job. *Occupational reinforcers* are defined as the environmental stimulus conditions (e.g., the Army's provision of opportunities to learn new skills) associated with persons' work values (Dawis & Lofquist, 1984). Thus, the WVI provides an assessment of respondents' work values via the importance they place on the occupational reinforcers that comprise the WVI. The majority of reinforcers that appear on the WVI were derived from Dawis and Lofquist's (1984) taxonomy of occupational reinforcers. The other reinforcers on the WVI were created specifically for Select21 based on a review of (a) the general literature on work values (e.g., Schwartz, 1994), (b) research on the values of American youth (Sackett & Mavor, 2002), (c) ARI's Army Values study (Ramsberger, Wetzel, Sipes, & Tiggle, 1999), and (d) the Select21 job analysis results. These new reinforcers were added to help round out the Dawis and Lofquist taxonomy for use in the Army context. Complete details on the development of the WVI are presented in Van Iddekinge, Putka et al. (2005).

The WVI has four parts and takes respondents roughly 15 to 20 minutes to complete. In the first part of the WVI, respondents sort the 28 reinforcers into four categories of varying importance. For example, respondents place their seven most important reinforcers in Category A and their seven least important reinforcers in Category D. Respondents then rank order the importance of the reinforcers within each category. After completing their rankings within each category, respondents are presented with the full list of reinforcers in the order they ranked them. Upon reviewing this list, they make a line through it—above the line are reinforcers they deem important to have on their ideal job, and below the line are reinforcers they deem unimportant to have on their ideal job.

Scoring

The WVI produces 28 work value scale scores, one for each occupational reinforcer comprising the WVI. The algorithm used to score the WVI scales parallels the algorithm used to score the Minnesota Importance Questionnaire (MIQ; Gay, Weiss, Hendel, Dawis, & Lofquist,

1971) and the Occupational Information Network (O*NET) Work Importance Profiler (WIP; McCloy et al., 1999). We subsequently refer to this algorithm as the MIQ/WIP algorithm.[38] The MIQ and WIP are very similar to the WVI in content and format in that both (a) draw heavily on Dawis and Lofquist'(1984) taxonomy of occupational reinforcers for content and (b) involve rank ordering of reinforcers and differentiating between important and unimportant reinforcers as a final step in the assessment process. Applying the MIQ/WIP algorithm to the WVI data yields 28 work value scale scores that are expressed in a *z*-score metric. Scale scores greater than 0 indicate a given reinforcer is important to the respondent, and scale scores less than 0 indicate a reinforcer is not important to the respondent. A key benefit of the MIQ/WIP scoring algorithm is its ability to provide a better approximation of persons' normative standing on each work value than would be possible based on rank-order information alone (Hicks, 1970). This result is achieved by using data from the final step in the WVI assessment (i.e., differentiating between important and unimportant reinforcers) to establish an individual zero-point on each value's importance scale. Establishing such a zero-point allows for more meaningful between-person comparisons because the ipsativity of the assessment is reduced (Gay et al., 1971).

Method

Sample

A total of 765 Soldiers completed the WVI during the concurrent validation data collections (Wave 1 = 597, Wave 2 = 168). We did, however, eliminate the responses of 33 Soldiers who test administrators flagged as having questionable WVI data or who had exhibited extremely unlikely patterns of responding. Thus, the final analysis sample comprised 732 Soldiers (Wave 1 = 570, Wave 2 = 162).

Validation Strategy

As noted in the previous chapter, a key element of any measure of P-E fit is how "environment-side" data (e.g., the extent to which the Army reinforces each of the 28 work values) are assessed and used in subsequent validation efforts (Kristof, 1996). The WVI, like other Select21 measures, is an assessment of person attributes (in this case, work values). It does not reflect the extent to which a person's work values are reinforced by the Army environment. In earlier Select21 data collections, 69 Army NCOs completed the Army Description Inventory (ADI), a measure designed to assess the degree to which the Army environment reinforces each of the 28 WVI work values for first-term Soldiers. The development, administration, and psychometric properties of the ADI are fully described in Van Iddekinge, Putka et al. (2005). We used mean NCO ratings from the ADI on each reinforcer as the environment-side "profile" when validating the WVI against the Select21 criteria.[39] Taken together, data from the WVI and ADI

[38] Details of this algorithm are presented in Appendix I of the measure development report (Knapp et al., 2005).
[39] In previous Select21 data collections, a far smaller group of NCOs ($N = 6$) completed a future-oriented version of the ADI—the Future Army Description Inventory (FADI). Although we initially considered creating fit measures based on comparison of the WVI and FADI, as was the case with the FAES in Chapter 12, preliminary analyses suggested that the results we would achieve using the FADI would be very similar to those achieved using the ADI (which is based on a far larger sample of NCOs). Thus, in this chapter the ADI served as the sole source of environment-side data.

can be combined to form an indirect, objective measure of P-E (Army) fit (Kristof-Brown, Zimmerman, & Johnson, 2005).

We adopted a validation strategy for the WVI that parallels the one we used for the WPS in the previous chapter. Specifically, we constructed four types of WVI composites that we subsequently validated against the Select21 criteria: (a) traditional profile similarity indexes (i.e., fit indexes), (b) regression weighted composites, (c) unit weighted composites, and (d) referent-based composites.[40] We discuss each of these in turn.

Traditional Profile Similarity Indexes

The first type of composites we constructed assess the similarity (or dissimilarity) between a Soldier's profile of scale scores on the WVI and the mean profile provided by NCOs on the ADI. As with the WPS, we calculated D^2 and Pearson *r* profile similarity indexes and estimated their criterion-related validity for predicting each of the Select21 criteria.

Regression Weighted Composites

We also used the approach described by Putka (2005) to create regression weighted WVI composites for this validation effort. One regression weighted composite was constructed for each Select21 criterion (i.e., we attempted to create optimal composites for each criterion).

Unit Weighted Composites

As we did for the WPS in the previous chapter, we also constructed unit weighted composites of WVI scales targeting each Select21 criterion. The process used to form these composites paralleled the process used to create the unit weighted WPS composites. Once the regression weighted composite targeting a given criterion was formed, we calculated zero-order correlations between the given criterion and each WVI scale that entered the final model for that criterion.[41] Only those WVI scales which had significant validities were included in the unit weighted composites for that criterion. All scales that that entered the unit weighted composite were given a weight of +1 or -1 (depending on the direction of their criterion-related validity).

Referent-Based Composites

In addition to the above composites, all of which have analogues to WPS composites described in the previous chapter, we also constructed a composite that arises naturally from the format of the WVI. Upon gathering ADI data from NCOs, we sorted occupational reinforcers into three categories, (a) those that are in high supply in the Army for first-term Soldiers, (b) those that are in moderate supply in the Army for first-term Soldiers, and (c) those that are in low

[40] We did not construct subjectively weighted composites. Examination of the criterion-related validities of the individual WVI scales comprising the unit weighted composites revealed that they varied to a far lesser extent compared to the WPS scales. As such, if we followed the strategy for constructing subjectively weighted composites outlined in Chapter 12, we would not have given any scale substantially higher or lower subjective weights (i.e., they would have all been unit weighted), and thus, any subjectively weighted composites we would have formed would not have differed from the unit weighted composites.

[41] Appendix I of Knapp et al. (2005) describes how the regression composites were formed (see also Putka, 2005).

supply in the Army for first-term Soldiers (Van Iddekinge, Putka et al., 2005). Based on these results, we constructed a simple "referent-based" WVI composite that reflected the proportion of times Soldiers' ranked reinforcers from the high supply category as more important than reinforcers from the low supply category.[42] The rationale behind constructing this composite and estimating its criterion-related validity stems from our hypothesis that Soldiers who prefer reinforcers that are in high supply in the Army over reinforcers that are in lower supply in the Army will have more positive attitudes towards the Army (or conversely, Soldiers who prefer reinforcers that are in low supply in the Army over reinforcers that are in high supply in the Army will have more negative attitudes towards the Army).

Cross-Validation

As was the case with the WPS composites, the various approaches to forming the WVI composites differ in terms of the degree to which their content and weighting are based on the sample data. As such, the criterion-related validities for some of these composites may reflect capitalization on chance more than others. For example, the content of the profile similarity indexes and referent-based WVI composite are not at all dependent on the sample data, as such, shrinkage is not an issue for these types of composites. On the other hand, the content and weighting of the weighted composites are, to a greater or lesser extent, derived from the sample data. For reasons cited in the previous chapter, we decided not to use formula-based estimates of cross-validity but instead to use Wave 1 and Wave 2 data to inform the extent to which these WVI composites might cross-validate.

As we did for the WPS, we present validation results based on WVI composites constructed on the full sample (Waves 1 and 2 combined). Basing these composites on the full sample allowed us to obtain the most stable estimates possible for the content and parameters of the weighted composites. After presenting these results, we show validity estimates for models based solely on the Wave 1 sample. We also show cross-validities for WVI composites in Wave 2 by taking the content and weighting underlying Wave 1 WVI composites and applying them to the Wave 2 data. Comparing the Wave 1 validities to the Wave 2 cross-validities allowed us to estimate the amount of shrinkage one might expect to observe from following the modeling processes we used to construct different types of WVI composites (e.g., regression, unit weighted). It is important to note that comparison of Wave 1 validities and Wave 2 cross-validities will only provide a *rough* estimate of how well the full sample WVI composites would be expected to cross-validate. First, all else being equal, the validity of the full sample WVI composites should be more stable than those based solely on Wave 1 data due to a larger sample size. Also, given that the full sample and Wave 1 sample only partially overlap, the content and weighing of the full sample and Wave 1 WVI composites may not be identical (even for those composites targeting the same criterion).

[42] Actually, the referent-based composite described here is just one example of a referent-based composite that could be formed based on the WVI data (e.g., another would be the proportion of times Soldiers rank low supply reinforcers over moderate supply reinforcers). Van Iddekinge, Putka et al. (2005) provided a more complete description of referent-based composites that can be created based on the WVI. The reason we limited our focus to this particular composite was that preliminary validation analyses indicated this composite held the most promise for predicting the Select21 performance and attitudinal criteria.

Results

Table 11.1 shows descriptive statistics for the WVI scale scores in the full sample.[43] On average, Soldiers most preferred work that provides opportunities for Advancement, Comfort, Achievement, and Leisure Time. Soldiers expressed least preference for work that provides opportunities for Travel, Influence, Activity, Team Orientation, and Independence. All of the WVI scales exhibited good levels of variability.

Table 11.1. Descriptive Statistics for WVI Scales

Scale	*M*	*SD*	Scale	*M*	*SD*
Ability Utilization	0.36	1.14	Independence	-0.59	1.33
Achievement	0.50	1.18	Influence	-0.78	1.06
Activity	-0.72	1.18	Leadership Opportunities	0.14	1.27
Advancement	0.87	1.14	Leisure Time	0.47	1.21
Autonomy	0.15	1.18	Personal Development	0.10	1.19
Comfort	0.64	1.24	Physical Development	-0.24	1.23
Co-Workers	-0.21	1.12	Recognition	-0.02	1.21
Creativity	-0.11	1.16	Social Service	-0.02	1.26
Emotional Development	-0.51	1.19	Social Status	0.43	1.26
Esteem	-0.43	1.17	Societal Contribution	-0.15	1.28
Feedback	-0.24	1.09	Supportive Supervision	0.16	1.27
Fixed Role	0.05	1.17	Team Orientation	-0.61	1.16
Flexible Schedule	0.02	1.22	Travel	-1.13	1.30
Home	-0.72	1.23	Variety	-0.13	1.14

Note. $n = 732$.

Table 11.2 shows raw zero-order intercorrelations among the WVI scales. On average, the WVI scales showed moderate levels of intercorrelation (mean $r = .46$). Interestingly, no negative correlations were observed. Often when dealing with forced choice measures such as the WVI, many intercorrelations are negative due to the ipsativity of the data (Hicks, 1970). These results were consistent with our contention that the WIP/MIQ algorithm reduces the ipsativity of the WVI scores, and in turn, enhances the degree to which the scores provide estimates of respondents' normative standing on each WVI scale.

In the field test, we found strong evidence for a six factor structure underlying the WVI scales that corresponded in meaningful ways to the factor structure underlying the MIQ and WIP interest measures (Van Iddekinge, Putka et al., 2005). For the present research, we attempted to replicate that structure, but were unable to do so. An exploratory factor analysis (EFA) of the data produced a four-factor solution which had several cross-loadings and factors that were difficult to interpret. The lack of simple structure for this sample may stem from differences between Soldiers in the field test sample and Soldiers in the concurrent validation sample.

[43] Given the partially-ipsative nature of the WVI no internal consistency estimates are provided for the WVI scales.

Table 11.2. Intercorrelations among WVI Scales

Scale		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Ability Utilization	.																										
2	Achievement	.59	.																									
3	Activity	.55	.53	.																								
4	Advancement	.55	.58	.49	.																							
5	Autonomy	.53	.50	.46	.53	.																						
6	Comfort	.46	.53	.39	.50	.49	.																					
7	Co-Workers	.55	.54	.51	.52	.45	.54	.																				
8	Creativity	.62	.54	.47	.45	.54	.52	.49	.																			
9	Emotional Development	.51	.44	.50	.46	.36	.34	.51	.38	.																		
10	Esteem	.55	.59	.49	.56	.47	.49	.58	.50	.55	.																	
11	Feedback	.60	.66	.56	.59	.49	.51	.53	.53	.49	.60	.																
12	Fixed Role	.51	.53	.52	.56	.43	.49	.51	.41	.48	.51	.60	.															
13	Flexible Schedule	.51	.41	.42	.46	.45	.61	.53	.48	.39	.47	.43	.41	.														
14	Home	.49	.48	.46	.48	.45	.50	.57	.49	.45	.57	.48	.48	.54	.													
15	Independence	.40	.39	.40	.31	.56	.42	.26	.51	.27	.34	.38	.38	.41	.36	.												
16	Influence	.55	.49	.54	.55	.49	.40	.50	.47	.59	.56	.51	.52	.47	.50	.37	.											
17	Leadership Opportunities	.48	.53	.42	.63	.43	.33	.47	.41	.49	.48	.62	.56	.29	.39	.23	.55	.										
18	Leisure Time	.49	.49	.38	.50	.51	.60	.47	.51	.35	.44	.46	.43	.56	.52	.44	.42	.33	.									
19	Personal Development	.61	.49	.54	.54	.42	.40	.57	.46	.63	.59	.54	.48	.47	.47	.27	.57	.49	.36	.								
20	Physical Development	.50	.47	.41	.50	.39	.32	.47	.35	.54	.42	.46	.48	.39	.38	.27	.47	.48	.39	.49	.							
21	Recognition	.49	.59	.44	.58	.44	.50	.53	.48	.42	.62	.60	.45	.47	.49	.37	.50	.46	.45	.47	.40	.						
22	Social Service	.49	.55	.46	.47	.38	.40	.51	.39	.46	.45	.53	.51	.32	.43	.24	.44	.53	.33	.46	.42	.38	.					
23	Social Status	.44	.56	.41	.56	.45	.47	.51	.36	.40	.49	.53	.48	.38	.45	.26	.42	.54	.42	.39	.44	.56	.51	.				
24	Societal Contribution	.49	.59	.45	.51	.40	.41	.48	.42	.44	.47	.52	.48	.33	.45	.25	.45	.50	.37	.43	.44	.41	.70	.54	.			
25	Supportive Supervision	.45	.52	.43	.57	.36	.52	.52	.36	.47	.50	.62	.59	.35	.40	.21	.47	.50	.38	.49	.46	.47	.49	.49	.46	.		
26	Team Orientation	.47	.45	.45	.46	.34	.43	.63	.42	.51	.52	.49	.44	.47	.49	.14	.55	.44	.39	.53	.43	.43	.50	.41	.46	.48	.	
27	Travel	.43	.41	.40	.40	.42	.34	.40	.42	.37	.36	.45	.36	.36	.21	.38	.40	.41	.37	.39	.43	.35	.33	.32	.38	.34	.35	.
28	Variety	.53	.53	.54	.52	.48	.49	.50	.51	.40	.44	.52	.50	.45	.42	.42	.44	.49	.48	.48	.46	.44	.46	.43	.46	.48	.43	.50

Note. $n = 732$. All correlations are statistically significant ($p < .05$, two-tailed).

For example, Soldiers in the field test sample were new recruits who had yet to be exposed to the Army environment; they completed the WVI immediately before entering basic training. On the other hand, Soldiers in the concurrent validation sample had generally been in the Army 18 to 36 months, and as such, completed the WVI well into their first-term of service. Based on Schneider's attraction-selection-attrition (ASA) hypothesis, one would expect the group of Soldiers in the concurrent validation sample to be more homogenous in terms of their work values than Soldiers in the field test sample (Schneider, 1987). This homogeneity may arise from the Army's training and socialization process, as well as attrition among Soldiers who enter the Army and find that they do not fit. The way this homogeneity may manifest itself in patterns of covariance among the WVI scales is that fewer factors may underlie the data. The reason for this pattern could be that a larger first factor (reflecting shared Soldier values) accounts for more of the covariation among work values. EFA of the concurrent validation data were consistent with this possibility in that a large first factor emerged from the data and it comprised several values that are reinforced by the Army, yet have historically loaded on different work value factors (e.g., Social Service, Feedback; Dawis & Lofquist, 1984).

Criterion-Related Validity Estimates

The previous section provided details on basic psychometric properties of the WVI scales. These scales (along with data from the ADI) provided the basis for the WVI composites discussed in this section. Table 11.3 shows criterion-related validity estimates for WVI composites in the full sample. [44] The table shows both uncorrected and corrected criterion related validity estimates for each of the 10 Select21 criteria. Analysis details are provided in Chapter 6. Criterion-related validity estimates for the "weighted" composites (i.e., regression and unit weighted composites) were not adjusted for shrinkage due to the issues summarized in Chapter 10. Later sections of this chapter will present validity estimates by sample to address the issue of how well the weighted composites cross-validate.

The results in Table 11.3 indicate the WVI has substantial promise as a predictor of the Select21 criteria, particularly the attitudinal criteria. Good levels of validity were also found for predicting the Achievement and Effort performance composite.

With regard to the magnitude of the criterion-related validity estimates, they were fairly impressive in both an absolute sense and in comparison to estimates in the literature. For example, in Project A, the average unadjusted multiple correlation among the three composites from the Job Orientation Blank (JOB) and Satisfaction with the Army (across MOS) was .11 in a longitudinal validation sample (Knapp & Carter, 2003, p. 48).[45] As shown in Table 11.3, the regression weighted WVI composite targeting Satisfaction with the Army had an uncorrected validity of .48. As with comparisons made to Project A results made in the WPS chapter, caution should be taken to not overinterpret these results, given the concurrent nature of the Select21 sample.

---

[44] Table 11.3 does not show criterion related-related validity estimates for regression and unit weighted composites for Teamwork because we did not "model" this criterion due to its unreliability (cf. Chapter 5).

[45] The JOB was a work-values measure developed for use in Project A and based on Dawis and Lofquist's (1984) Theory of Work Adjustment.

Table 11.3. Criterion-Related Validity Estimates for WVI Composites in the Full Sample

	Performance Criteria					Attitudinal Criteria				
Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
D^2 Fit Index	**-.08**	-.03	.02	-.03	-.06	**-.10**	**-.12**	**.10**	-.05	.00
Pearson *r* Fit Index	.05	**.20**	**.13**	**.08**	**.09**	**.37**	**.39**	**-.24**	**.28**	**.18**
Referent-Based	.06	**.25**	**.14**	**.09**	**.11**	**.38**	**.39**	**-.26**	**.28**	**.17**
Regression	**.22**	**.30**	**.29**	.	**.18**	**.48**	**.50**	**.36**	**.39**	**.30**
Unit	**.14**	**.28**	**.24**	.	**.14**	**.47**	**.46**	**-.34**	**.38**	**.29**
Unit AE	**.10**	**.28**	**.15**	**.13**	**.10**	**.34**	**.38**	**-.25**	**.27**	**.19**
Unit ASat	**.07**	**.24**	**.19**	**.03**	**.11**	**.47**	**.47**	**-.32**	**.37**	**.27**
	Corrected Validity Estimates									
D^2 Fit Index	-.16	-.07	.02	-.07	-.12	-.11	-.14	.14	-.04	.00
Pearson *r* Fit Index	-.03	.16	.13	.09	.06	.39	.42	-.25	.30	.20
Referent-Based	-.01	.23	.14	.12	.08	.40	.43	-.28	.30	.19
Regression	.30	.32	.30		.26	.51	.55	.41	.41	.32
Unit	.21	.28	.24		.19	.50	.51	-.39	.40	.31
Unit AE	.07	.28	.15	.19	.10	.37	.43	-.29	.28	.20
Unit ASat	.02	.23	.19	.03	.10	.50	.51	-.36	.39	.29

Note. n = 525 (AE criterion), n = 700 (all other performance criteria), n = 663-680 (attitudinal criteria). Referent = Referent-based composite score reflecting proportion of times Soldiers ranked high supply WVI reinforcers over low supply WVI reinforcers. Regression = Regression-weighted composite score specific to each criterion optimized in the full sample. Unit = Unit-weighted composite score specific to each criterion based on regression analyses in the full sample. Unit AE = Unit weighted composite score specific to the AE performance criterion. Unit ASat = Unit weighted composite score specific to the ASat attitudinal criterion. Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

While comparing favorably to past Army research, these results also compared favorably to past research in the civilian P-E fit literature. Past meta-analytic estimates of the criterion-related validity of indirect, objective measures of P-E fit for predicting satisfaction and intentions to quit (similar to attrition cognitions) were .29 and -.19, respectively (Kristof-Brown et al., 2005). With the exception of the D^2 fit index, the validity of all other WVI composites exceeded these meta-analytic estimates for the aforementioned criteria.

As was the case with the WPS composites, the regression weighted and unit weighted WVI composites exhibited notably higher levels of validity (about .10 higher on average for the attitudinal criteria) compared to the Pearson *r* fit index. Furthermore, once again, very little validity was lost by using unit weights as opposed to regression weights. The criterion-related validity of the referent based composite was very comparable to the criterion-related validity of the Pearson *r* fit index. Similar to findings from the previous chapter, these results provide further evidence that profile similarity indexes such as D^2 and Pearson *r* commonly used in the P-E fit literature artificially constrain observed person-environment-criterion relations.

As with the WPS composites, we were able to obtain good levels of validity by using composites optimized on one criterion as predictors of other criteria. For example, the unit weighted composite targeting Satisfaction with the Army had criterion-related validities for predicting all other attitudinal criteria that exceeded .26 in magnitude. Given these results and the desirability of having a parsimonious set of WVI predictors, we limited our attention to only two of the 21 composites summarized in Table 11.3 for subsequent cross-instrument analyses in this report (see Chapters 13-15), namely the Unit Achievement and Effort and Unit Satisfaction with the Army composites. Of the WVI composites, these two had the highest absolute validity (on average) for predicting the performance and attitudinal criteria, respectively.

Composition of WVI Composites

Table 11.4 shows the composition of the weighted WVI composites. A primary difference between the regression-weighted WPS composites and the regression weighted WVI composites is that more evidence for non-linearity in WVI-criterion relationships emerged. This is evidenced by the non-linear functions of WVI-ADI scores (e.g. absolute WVI-ADI difference scores, spline adjustment terms) that entered the prediction model for various criteria. While the inclusion of these terms suggests the importance of adopting a regression based approach to building P-E fit composites (e.g., Edwards, 1993; Putka, 2005), their importance is greatly tempered by the fact that the unit weighted WVI composites (which contain no non-linear terms, and are based solely on WVI data) achieved comparable levels of criterion-related validity to their regression weighted counterparts (see Table 11.3).

Table 11.4. Composition of WVI Composites

	Performance Criteria				Attitudinal Criteria				
Scale	GTP	AE	PF	FXP	ASat	AFit	ACog	CInt	FAA
Ability Utilization	.	.	.	.	0.13[a]	.	.	.	.
Achievement	.	.	.	.	.	.	-0.11[a]	.	.
Activity	.	.	.	.	.	.	.	.	.
Advancement	.	.	.	.	0.08[a]	.	.	0.10[a]	.
Autonomy	0.13[a]	.	.	.	.	.	.	.	.
Comfort	.	.	.	.	-0.16[a]	-0.21[a]	0.17[a]	-0.12[a]	-0.15[a]
Co-Workers	.	.	.	.	.	.	.	.	.
Creativity	.	.	.	.	-0.16[a]	-0.12[a]	.	.	.
Emotional Development	.	0.11[a]	.	.	0.18[a]	0.12[a]	-0.12[a]	0.03[a]	.
Emotional Development (DK75L)	.	.	.	.	.	.	.	0.11	.
Emotional Development (DK75U)	.	.	.	.	.	.	.	-0.05	.
Esteem	.	.	.	.	.	.	.	.	.
Feedback	.	.	.	.	.	.	.	.	.
Fixed Role	.	.	-0.16[a]	.	.	0.08[a]	.	.	.
Flexible Schedule	.	.	.	.	-0.12[a]	.	.	-0.12[a]	.
Home	.	.	.	.	.	.	.	.	.
Independence	0.03	-0.14[a]	-0.10[a]	.	-0.14[a]	-0.20[a]	0.14[a]	-0.10[a]	-0.10[a]
Independence (QSK)	-0.18	.	.	.	.	.	.	.	.
Independence (SD)	.	.	.	-0.09	.	.	.	.	.
Influence	.	.	.	.	.	.	.	.	.
Leadership Opportunities	0.10[a]	0.13[a]	.	.	.	0.27[a]	.	0.10[a]	0.13[a]
Leadership Opportunities (QSK)	.	.	.	.	.	-0.13	.	.	.

Table 11.4. (Continued)

	Performance Criteria				Attitudinal Criteria				
Scale	GTP	AE	PF	FXP	ASat	AFit	ACog	CInt	FAA
Leisure Time	0.15	.[a]	-0.10[a]	.	-0.17[a]	-0.13[a]	0.13[a]	-0.15[a]	-0.11[a]
Leisure Time (AD)	.	-0.12	.	.	.	.	.	.	.
Leisure Time (LSK)	-0.18	.	.	.	.	.	.	.	.
Personal Development	.	.	.	.	.	.	.	.	.
Personal Development (AD)	.	.	.	-0.07	.	.	.	.	.
Physical Development	.	.	0.29[a]	.	0.20[a]	0.20[a]	-0.22[a]	0.14[a]	.
Physical Development (AD)	.	.	.	.	.	.	.	.	-0.13[a]
Recognition	.	.	.	.	.	.	.	.	.
Social Service	.	.	.	.	.	.	.	.	.
Social Status	.	.	.	.	0.10[a]	0.10[a]	.	.	.
Societal Contribution	.	-0.03[a]	.	.	.	.	.	.	.
Societal Contribution (QSK)	.	0.17	.	.	.	.	.	.	.
Supportive Supervision	.	.	.	.	.	.	.	.	.
Team Orientation	.	.	.	0.11[a]	.	.	.	.	.
Travel	.	.	.	.	0.09[a]	0.09[a]	.[a]	0.13[a]	0.13[a]
Travel (AD)	.	.	.	.	.	.	0.07	.	.
Variety	.	.	.	.	.	.	.	.	.

Note. Cell values reflect standardized beta weights for the WVI regression-based composite targeting the given criterion. If no cell value is listed for a given WVI scale, then it means that the WVI scale was not included in the composite for the given criterion. All scales that have superscripts on their standardized beta weights were included in unit-weighted composites targeting the given criterion and received a weight of +1 or -1 (depending on the direction of its zero-order correlation with the criterion). Scales with parenthetical notations following them had non-linear relationships with the given criterion. For those scales, the non-linear terms entered into the model were as follows: AD = Absolute difference between the given WVI scale and corresponding ADI scale; SD = Squared difference between the given WVI scale and corresponding ADI scale; LSK = Linear spline adjustment term modeling a knot at the mean ADI value for the given WVI scale; QSK = Quadratic spline adjustment term modeling a knot at the mean ADI value for the given WVI scale; DK75L = Linear spline adjustment term modeling a knot 0.75 points below the mean ADI value for the given WVI scale; DK75U = Linear spline adjustment term modeling a knot 0.75 points above the mean ADI value for the given WVI scale. For further details on spline adjustment terms, see Putka (2005). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Like the composition of the WPS composites targeting the Select21 performance criteria, there was little consistency in the composition of WVI composites designed to predict the different performance criteria. On the attitudinal side, there was far more consistency in the composition of the WVI composites. For example, the WVI Comfort, Leisure Time, Independence, and Travel scales played a role in the weighted composites for all five attitudinal criteria. Furthermore, WVI Emotional Development and Physical Development played a role in the weighted composites for four of the five attitudinal criteria (all except Future Army Affect). The fact that these characteristics consistently emerged across criteria (both in magnitude and direction) appears consistent with the extent to which those work values are reinforced by the Army environment. For example, the Army generally offers Soldiers opportunities for travel, emotional development, and physical development, but the Army arguably offers less opportunities for comfort, leisure time, and independence (at least for first-term Soldiers). Like the findings regarding the WPS Physical facet presented in the previous chapter, physical fitness

related content (i.e., valuing opportunities for physical development) once again appeared to have a key role in predicting attitudinal criteria. As mentioned before, such results are consistent with past research which has suggested physical fitness plays a key role in understanding the attitudes and behaviors of Soldiers (Strickland, 2005).

Relations among Composites

The criterion-related validity estimates of many WVI composites were presented in Table 11.3. Table 11.5 shows the correlation between the final two WVI composites we chose to move forward with and the other WVI composites. Not surprisingly, the two final composites were highly related to the other weighted composites that targeted the same criterion (e.g., the unit weighted composite targeting Satisfaction with the Army was correlated .97 with the regression weighted composite targeting Satisfaction with the Army). In general, both of the final composites were moderately to strongly related to the other composites, with many correlations exceeding .60. This was particularly true for relations between the unit weighted composite targeting Satisfaction with the Army, and composites targeting the other attitudinal criteria (all but one of these correlations exceeded .80 in magnitude). This finding is not surprising given the

Table 11.5. Correlations between Final WVI Composites and Other WVI Composites

	Final WVI Composites	
All WVI Composites	Unit AE	Unit ASat
1. D^2 Fit Index	-.25	-.26
2. Pearson *r* Fit Index	.65	.73
3. Referent-Based	.65	.74
4. Regression General Technical Proficiency	.52	.38
5. Unit General Technical Proficiency	.41	.27
6. Regression Achievement and Effort	.83	.60
7. Unit Achievement and Effort	1.00	.73
8. Regression Physical Fitness	.35	.55
9. Unit Physical Fitness	.27	.38
10. Regression Future Expected Performance	.45	.31
11. Unit Future Expected Performance	.39	.22
12. Regression Satisfaction with the Army	.70	.97
13. Unit Satisfaction with the Army	.73	1.00
14. Regression Perceived Army Fit	.77	.90
15. Unit Perceived Army Fit	.80	.90
16. Regression Attrition Cognitions	-.69	-.85
17. Unit Attrition Cognitions	.73	.87
18. Regression Career Intentions	.73	.90
19. Unit Career Intentions	.78	.92
20. Regression Future Army Affect	.64	.79
21. Unit Future Army Affect	.69	.83

Note. $n = 732$. Correlations that appear in boxes are for those WVI composites that target the same criterion as the WVI composite shown at the top of the given column. All correlations are statistically significant ($p < .05$, one-tailed). ASat = Satisfaction with the Army. AE= Achievement and Effort.

moderate to high correlations observed between the attitudinal criteria in Chapter 3. In contrast to relations between the final WPS composites and the Pearson *r* fit index, both of the final WVI composites were strongly related to the Pearson *r* fit index (Unit Achievement and Effort: $r = .65$, Unit Satisfaction with the Army: $r = .73$). Such findings indicate that these composites shared a substantial amount of variance with an index of the similarity between Soldiers' profiles on the WVI and the Army profile based on the ADI.

Cross-Validation of Composites

Table 11.6 shows criterion-related validity estimates for WPS composites in the Wave 1 and Wave 2 samples.[46] Unlike Table 11.3, the weighted WVI composites in this table were constructed based on the Wave 1 data only. Thus, the Wave 2 validity estimates represent cross-validities (i.e., criterion-related validities based on applying Wave 1 parameters to Wave 2 data). Based on Table 11.6, the weighted WVI composites targeting the attitudinal criteria appeared to retain their validity better than did the weighted WVI composites targeting the performance criteria. Unfortunately, for both sets of weighted composites, the estimated validities appeared to take a substantial hit upon cross-validation in the Wave 2 sample. On average, regression weighted composites targeting performance criteria lost 40.4% of their validity, whereas the unit weighed composites targeting performance criteria lost 63.8% of their validity (based on comparison of corrected Wave 1 and Wave 2 validity estimates). On average, regression weighted composites targeting attitudinal criteria lost 29.5% of their validity, whereas unit weighed composites targeting attitudinal criteria lost 27% of their validity (again based on comparison of corrected validity estimates). Despite the losses in validity, the regression and unit weighted composites still had good levels of validity for predicting all attitudinal criteria in the Wave 2 sample (except Future Army Affect). Nevertheless, compared to the full sample results, the cross-validated criterion-related validity estimates for the weighted WVI composites now appear far more similar to the validity achieved by using the Pearson *r* fit index. Indeed, for predicting Attrition Cognitions and Career Intentions, the Pearson *r* fit index actually exhibited corrected validity estimates that were about .08 to .16 higher than those obtained for the regression weighted and unit weighted WVI composites targeted at predicting those criteria.

[46] Note, unlike Table 11.3, this table does not show the validity of the Unit AE or Unit ASat composites for predicting all other criterion composites. Remember the purpose of this table is not to cross-validate the actual composites formed on the full sample, but rather to gain insight into how well the process used to create those composites results in scores whose validity holds up upon cross-validation. The reason why Unit AE and Unit ASat were highlighted in Table 11.3 was because those were the two final WVI composites that would be used in subsequent chapters in this report.

Table 11.6. Criterion-Related Validity Estimates for WVI Composites in the Wave 1 and Wave 2 Samples

	Performance Criteria					Attitudinal Criteria				
Sample/Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Validity Estimates									
Wave 1 Sample										
D^2 Fit Index	**-.10**	-.07	.04	**-.08**	**-.07**	**-.14**	**-.17**	**.12**	-.05	-.05
Pearson *r* Fit Index	.02	**.16**	**.15**	.06	.06	**.36**	**.37**	**-.21**	**.24**	**.21**
Referent-Based	.02	**.21**	**.16**	.07	**.08**	**.38**	**.38**	**-.24**	**.25**	**.20**
Regression (W1)	**.27**	**.35**	**.33**	.	**.18**	**.51**	**.49**	**.32**	**.36**	**.35**
Unit (W1)	**.16**	**.24**	**.27**	.	**.15**	**.48**	**.41**	**-.30**	**.34**	**.32**
Wave 2 Sample										
D^2 Fit Index	-.01	.04	-.05	**.16**	.02	.02	.01	.03	-.05	.10
Pearson *r* Fit Index	**.14**	**.27**	.01	.11	**.17**	**.37**	**.43**	**-.34**	**.41**	.12
Referent-Based	**.19**	**.34**	.01	.12	**.17**	**.36**	**.42**	**-.32**	**.35**	.12
Regression (W1)	.07	**.18**	**.14**	.	.10	**.38**	**.44**	**.27**	**.29**	**.14**
Unit (W1)	-.03	.12	.12	.	.02	**.39**	**.32**	**-.21**	**.32**	**.23**
	Corrected Validity Estimates									
Wave 1 Sample										
D^2 Fit Index	-.16	-.10	.04	-.15	-.13	-.15	-.19	.16	-.05	-.05
Pearson *r* Fit Index	-.08	.12	.15	.06	.01	.38	.39	-.20	.25	.22
Referent-Based	-.08	.17	.16	.07	.03	.40	.40	-.24	.26	.21
Regression (W1)	.37	.38	.33	.	.28	.54	.54	.39	.37	.38
Unit (W1)	.24	.26	.26	.	.24	.51	.46	-.37	.35	.34
Wave 2 Sample										
D^2 Fit Index	-.13	-.04	-.05	.23	-.08	.03	.06	.03	.03	.14
Pearson *r* Fit Index	.14	.27	.02	.17	.24	.39	.48	-.40	.43	.14
Referent-Based	.24	.35	.01	.19	.26	.38	.46	-.37	.36	.13
Regression (W1)	.20	.23	.15	.	.22	.40	.48	.32	.32	.14
Unit (W1)	.01	.16	.13	.	.07	.41	.35	-.24	.32	.24

Note. n_{Wave1} = 385 (AE criterion), n_{Wave1} = 547 (all other performance criteria), n_{Wave1} = 506-523 (attitudinal criteria). n_{Wave2} = 140 (AE criterion), n_{Wave2} = 153 (all other Performance criteria), n_{Wave2} = 157 (attitudinal criteria). Regression (W1) = Regression-weighted composite score specific to each criterion optimized in the Wave 1 sample. Unit (W1) = Unit-weighted composite score specific to each criterion based on regression analyses in Wave 1 sample. Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Incremental Validity Estimates

In the previous section, we provided evidence for the criterion-related validity of the WVI. Here we focus on the degree to which it increments the validity of the AFQT. Table 11.7 shows incremental validity estimates for the WVI composites in the full concurrent validation sample. The estimates presented in Table 11.7 show that the WVI had a substantial level of incremental validity over the AFQT for predicting the attitudinal criteria. This finding is not surprising given the general lack of validity of the AFQT for predicting attitudinal criteria, and the good validity of the WVI for predicting attitudinal as shown above in Table 11.3. With regard to the performance criteria, the incremental validity of the WPS composites over the AFQT was notable for the Achievement and Effort and Physical Fitness performance composites, but not General Technical Proficiency. This finding was consistent with our expectations.

Table 11.7. Incremental Validity Estimates for WVI Composites in the Full Sample

	Performance Criteria					Attitudinal Criteria				
Composite	GTP	AE	PF	TEAM	FXP	ASat	AFit	ACog	CInt	FAA
	Uncorrected Incremental Validity Estimates									
AFQT	**.31**	**.17**	.01	**.08**	**.19**	.00	.00	**-.10**	**-.07**	-.03
D^2 Fit Index	.00	.00	.01	.00	.00	**.10**	**.12**	**.03**	.02	.00
Pearson *r* Fit Index	**.02**	**.12**	**.12**	**.04**	**.04**	**.37**	**.39**	**.18**	**.22**	**.15**
Referent-Based: High over Low	**.02**	**.16**	**.13**	**.05**	**.04**	**.38**	**.39**	**.20**	**.22**	**.14**
Regression Weights (F)	**.05**	**.18**	**.29**	.	**.06**	**.48**	**.50**	**.28**	**.32**	**.26**
Unit Weights (F)	**.02**	**.17**	**.24**	.	**.04**	**.47**	**.47**	**.26**	**.32**	**.26**
Unit Weight AE (F)	**.03**	**.17**	**.14**	**.08**	**.03**	**.34**	**.38**	**.18**	**.20**	**.16**
Unit Weight ASat (F)	**.02**	**.14**	**.18**	.01	**.04**	**.47**	**.47**	**.25**	**.31**	**.24**
	Corrected Incremental Validity Estimates									
AFQT	.54	.30	.02	.20	.38	-.01	.01	-.19	-.11	-.06
D^2 Fit Index	.00	.00	.00	.00	.00	.07	.11	.01	.00	.00
Pearson *r* Fit Index	.01	.08	.10	.04	.03	.38	.43	.17	.19	.13
Referent-Based: High over Low	.01	.12	.11	.05	.03	.40	.43	.19	.19	.11
Regression Weights (F)	.03	.14	.29	.	.05	.51	.55	.29	.30	.25
Unit Weights (F)	.01	.14	.23	.	.03	.49	.51	.27	.29	.25
Unit Weight AE (F)	.02	.14	.12	.09	.02	.35	.42	.17	.17	.13
Unit Weight ASat (F)	.01	.11	.17	.00	.03	.49	.51	.25	.28	.22

Note. n = 503 (AE criterion), n = 675 (all other performance criteria), n = 636-656 (attitudinal criteria). Cell values for the AFQT represent zero-order correlations between the AFQT and the given criterion (shown for reference). Uncorrected incremental estimates reflect the difference between the Multiple *R* obtained when regressing the criterion on both the given composite and AFQT versus the *R* obtained when regressing the criterion only on the AFQT. Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Statistically significant incremental validities are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Subgroup Differences

Tables 11.8 and 11.9 show mean final WVI composite scores by gender and race/ethnicity, respectively. Though four statistically significant differences were found, the magnitudes of these effects sizes were generally small to moderate in magnitude (0.21 to 0.47), and in all cases, minority groups (e.g., females, Blacks, Hispanics) scored higher than the majority group (e.g., males, whites).

Table 11.8. Final WVI Composite Scores by Gender

		Male		Female	
WVI Composite	d_{FM}	*M*	*SD*	*M*	*SD*
Unit AE	**0.47**	-0.11	0.56	0.16	0.55
Unit ASat	0.09	-0.06	0.37	-0.03	0.35

Note. n_{Male} = 655. n_{Female} = 76. d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of males – mean of females)/*SD* of males. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 11.9. Final WVI Composite Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
WVI Composite	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unit AE	**0.21**	**0.37**	-0.12	0.57	0.00	0.58	-0.16	0.57	0.05	0.55
Unit ASat	-0.01	**0.24**	-0.07	0.38	-0.07	0.31	-0.08	0.38	0.01	0.36

Note. n_{White} = 522. n_{Black} = 141. $n_{White\ Non\text{-}Hispanic}$ = 407. $n_{Hispanic}$ = 142. d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites)/*SD* of Whites. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Differential Prediction

Tables 11.10 through 11.12 present the results of differential prediction analyses for the final WVI composites. Table 11.10 shows results for gender, Table 11.11 for race, and Table 11.12 for race/ethnicity.[47] Overall, the results indicate some evidence of intercept bias and differential prediction (i.e., slope bias) depending on the criterion, WVI composite, and demographic variable considered. In light of these findings, we discuss results from each of the tables in turn. We focus only on interpreting results for the criteria each WVI composite was meant to predict (Unit Achievement and Effort [Unit AE]—performance criteria; Unit Satisfaction with the Army [Unit ASat]—attitudinal criteria).

Table 11.10 reveals evidence of intercept bias for the Unit AE and Unit ASat composites by gender. Intercept bias was apparent when using Unit AE to predict Achievement and Effort, Physical Fitness, and Future Expected Performance, and when using Unit ASat to predict Satisfaction with the Army, Attrition Cognitions, and Future Army Affect. In the case of the Unit AE composite, women had Achievement and Effort and Future Expected Performance scores that were roughly 0.21 and 0.24 points higher than men (at mean levels of the Unit AE composite), and Physical Fitness scores that were roughly 0.25 points lower than men. These

[47] All WVI composite scores were standardized prior to conducting these analyses to ease interpretation of the unstandardized regression weights presented in these tables.

findings suggest that if computed by a prediction equation based on all Soldiers, Unit AE composite scores would tend to underpredict females' performance on Achievement and Effort and Future Expected Performance, and overpredict their performance on Physical Fitness. In the case of the Unit ASat composite, women had Satisfaction with the Army and Future Army Affect scores that were roughly 0.23 and 0.35 points lower than men (at mean levels of the Unit ASat composite), and Attrition Cognition scores that were roughly 0.29 points higher than men. These findings suggest that if computed by a prediction equation based on all Soldiers, Unit ASat composite cores would tend to overpredict women's Satisfaction with the Army and Future Army Affect, and underpredict their Attrition Cognitions.

Table 11.10. Differential Prediction Results for Final WVI Composites by Gender

	Unit AE WVI Composite					Unit ASat WVI Composite				
	Gender	WVI *b*		*r* by Gender		Gender	WVI *b*		*r* by Gender	
Criterion	*b*	M	F	M	F	*b*	M	F	M	F
Performance Criteria										
General Technical Proficiency	-0.04	0.05	0.08	**.10**	.14	-0.01	0.03	0.10	.05	.19
Achievement and Effort	**0.24**	0.15	0.05	**.29**	.09	**0.26**	0.13	0.06	**.26**	.12
Physical Fitness	**-0.25**	**0.10**	**0.31**	**.13**	**.36**	-0.16	**0.12**	**0.32**	**.17**	**.39**
Teamwork	0.12	0.05	0.20	**.09**	**.31**	**0.19**	0.00	0.14	.00	.22
Future Expected Performance	**0.21**	0.06	0.04	**.09**	.06	**0.22**	0.06	0.11	**.10**	.17
Attitudinal Criteria										
Satisfaction with the Army	**-0.29**	0.29	0.21	**.37**	**.28**	**-0.23**	0.36	0.41	**.46**	**.55**
Perceived Army Fit	-0.15	0.32	0.29	**.39**	**.34**	-0.05	0.37	0.43	**.46**	**.49**
Attrition Cognitions	**0.39**	-0.26	-0.27	**-.27**	**-.26**	**0.29**	-0.32	-0.28	**-.33**	**-.27**
Career Intentions	-0.13	0.29	0.35	**.26**	**.30**	-0.01	0.40	0.53	**.36**	**.44**
Future Army Affect	**-0.38**	0.20	0.15	**.21**	.16	**-0.35**	0.25	0.29	**.27**	**.30**

Note. $n_{Regression}$ = 524-699. n_{Male} = 460-631. n_{Female} = 64-73. Gender *b* = Unstandardized regression weight for gender (0 = male, 1 = female). WVI *b* = Unstandardized regression weight for the given WVI composite for males and females. *r* by Gender = Correlation between the given WVI composite and the given criterion for each gender. Regression weights for males and females are bolded if the WVI-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 11.10 reveals no evidence of slope bias for the Unit ASat composite but does indicate slope bias for the Unit AE composite when predicting Physical Fitness. Specifically, the Unit AE composite was more predictive of Physical Fitness for females ($b = 0.31$, $r = .36$) than for males ($b = 0.10$, $r = .13$).

Table 11.11 reveals that intercept bias by race (Black vs. white) was apparent when using Unit AE to predict General Technical Proficiency, Achievement and Effort, and Future Expected Performance, and when using Unit ASat to predict Attrition Cognitions. In the case of the Unit AE composite, Black Soldiers had General Technical Proficiency, Achievement and Effort, and Future Expected Performance scores that were roughly 0.16 and 0.26 points lower than White Soldiers (at mean levels of the Unit AE composite). These findings suggest that if computed by a prediction equation based on all Soldiers, Unit AE composite scores would tend to overpredict Black Soldiers' performance on these composites. In the case of the Unit ASat composite, Black Soldiers had Attrition Cognition scores that were roughly 0.35 points higher than White Soldiers

(at mean levels of the Unit ASat composite). These findings suggest that if computed by a prediction equation based on all Soldiers, Unit ASat composite scores would tend to underpredict Black Soldiers' Attrition Cognitions.

Table 11.11. Differential Prediction Results for Final WVI Composites by Race

	Unit AE WVI Composite					Unit ASat WVI Composite				
	Race	WVI *b*		*r* by Race		Race	WVI *b*		*r* by Race	
Criterion	*b*	W	B	W	B	*b*	W	B	W	B
Performance Criteria										
General Technical Proficiency	**-0.26**	**0.04**	**0.15**	.08	**.33**	**-0.23**	0.03	0.13	.05	**.24**
Achievement and Effort	**-0.16**	0.14	0.20	**.28**	**.37**	**-0.13**	0.13	0.19	**.26**	**.29**
Physical Fitness	0.00	0.13	0.07	**.17**	.09	0.02	0.15	0.13	**.21**	.15
Teamwork	-0.02	0.06	0.15	**.10**	**.24**	0.00	0.00	0.12	.01	.17
Future Expected Performance	**-0.16**	0.08	0.10	**.11**	**.18**	**-0.14**	0.07	0.18	**.11**	**.28**
Attitudinal Criteria										
Satisfaction with the Army	-0.10	0.28	0.26	**.36**	**.34**	-0.05	0.37	0.35	**.50**	**.38**
Perceived Army Fit	**-0.19**	0.34	0.29	**.42**	**.37**	-0.13	0.39	0.38	**.49**	**.41**
Attrition Cognitions	**0.41**	-0.29	-0.27	**-.30**	**-.29**	**0.35**	-0.34	-0.32	**-.36**	**-.29**
Career Intentions	-0.01	**0.36**	**0.15**	**.32**	.14	0.03	**0.47**	**0.18**	**.43**	.14
Future Army Affect	**-0.21**	0.17	0.30	**.18**	**.32**	-0.17	0.26	0.27	**.29**	**.24**

Note. $n_{\text{Regression}}$ = 479-634. n_{White} = 375-502. n_{Black} = 104-132. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black). WVI *b* = Unstandardized regression weight for the given WVI composite for Whites and Blacks. *r* by Race = Correlation between the given WVI composite and the given criterion for each race. Regression weights for Whites and Blacks are bolded if the WVI-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 11.11 also reveals evidence of slope bias for the Unit AE composite when predicting General Technical Proficiency, and for the Unit ASat composite when predicting Career Intentions. Specifically, the Unit AE composite was more predictive of General Technical Proficiency for Black Soldiers ($b = 0.15$, $r = .33$) than for White Soldiers ($b = 0.04$, $r = .08$), and the Unit ASat composite was more predictive of Career Intentions for White Soldiers ($b = 0.47$, $r = .43$) than for Black Soldiers ($b = 0.18$, $r = .14$).

Table 11.12 reveals no evidence of intercept bias or slope bias for the Unit AE composite by ethnicity (Hispanic vs. White non-Hispanic). Nevertheless, evidence of intercept bias was found for the Unit ASat composite when predicting Future Army Affect, and evidence of slope bias was found for the Unit ASat composite when predicting Attrition Cognitions. With regard to intercept bias, Hispanic Soldiers had Future Army Affect scores that were roughly 0.20 points higher than White non-Hispanic Soldiers (at mean levels of the Unit ASat composite). These findings suggest that if computed by a prediction equation based on all Soldiers, Unit ASat composite scores would tend to underpredict Hispanic Soldiers' performance on Future Army Affect. With regard to the slope bias, the Unit ASat composite was more predictive of Attrition Cognitions for White non-Hispanic Soldiers ($b = -0.38$, $r = -.41$) than for Hispanic Soldiers ($b = -0.18$, $r = -.19$).

Discussion

Based on the results presented in this chapter, the WVI appears to have substantial promise for predicting attitudinal criteria often found to be key precursors of attrition and re-enlistment behavior (Strickland, 2005). Results also indicate that the WVI has promise for predicting Achievement and Effort and Physical Fitness performance above and beyond the AFQT. The findings with regard to the criterion-related validity of the WVI observed in this chapter are generally stronger than those found in past Army research, as well as civilian research on P-E fit measures. Part of the reason for the success of the WVI in Select21 may be the more rigorous approach taken to modeling person-environment-criterion relations than has been typically reported in the research literature. In addition to exhibiting good levels of criterion-related validity, the final WVI composites recommended for future use exhibit only small to moderate group differences across genders and racial/ethnic groups. Further, in cases where such differences were found, they were in favor of the minority groups.

Table 11.12. Differential Prediction Results for Final WVI Composites by Ethnic Group

	Unit AE WVI Composite					Unit ASat WVI Composite				
	Eth	WVI *b*		*r* by Eth		Eth	WVI *b*		*r* by Eth	
Criterion	*b*	W	H	W	H	*b*	W	H	W	H
Performance Criteria										
General Technical Proficiency	-0.05	0.05	0.03	.08	.07	-0.04	0.03	0.01	.07	.02
Achievement and Effort	0.01	0.14	0.19	**.26**	**.35**	0.02	0.13	0.14	**.26**	**.27**
Physical Fitness	0.00	0.12	0.15	**.16**	**.19**	0.02	0.17	0.11	**.23**	.15
Teamwork	0.10	0.03	0.12	.05	**.22**	**0.12**	-0.01	0.05	-.02	.09
Future Expected Performance	0.04	0.07	0.06	**.10**	.09	0.04	0.07	0.05	**.11**	.08
Attitudinal Criteria										
Satisfaction with the Army	0.03	0.29	0.26	**.37**	**.33**	0.04	0.37	0.37	**.50**	**.49**
Perceived Army Fit	0.01	0.35	0.30	**.43**	**.37**	0.03	0.38	0.38	**.48**	**.49**
Attrition Cognitions	0.02	**-0.36**	**-0.03**	**-.37**	-.04	0.03	**-0.38**	**-0.18**	**-.41**	**-.19**
Career Intentions	-0.03	0.42	0.20	**.36**	**.18**	-0.04	0.49	0.38	**.44**	**.35**
Future Army Affect	**0.23**	**0.21**	**-0.07**	**.23**	-.08	**0.20**	0.27	0.15	**.31**	.16

Note. $n_{Regression}$ = 390-527. $n_{White\ non\text{-}Hispanic}$ = 294-390. $n_{Hispanic}$ = 96-137. Eth b = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). WVI *b* = Unstandardized regression weight for the given WVI composite for White non-Hispanics and Hispanics. *r* by Eth = Correlation between the given WVI composite and the given criterion for each ethnic group. Regression weights for White non-Hispanics and Hispanics are bolded if the WVI-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

While the aforementioned results are promising, there are some causes for concern with the WVI. Specifically, the final set of analyses revealed some evidence that predictive biases may result from using the WVI in selection contexts. Some cases of bias, such as the intercept differences found across genders, seemed to be due primarily to the subgroup differences on the criteria of interest rather than to the WVI itself (see Chapters 3 through 5). In other cases, the observed biases may be more problematic. For example, we found the Unit Satisfaction with the Army WVI composite was more predictive of Career Intentions for White Soldiers than for Black Soldiers and also more predictive of Attrition Cognitions for White non-Hispanic Soldiers than for Hispanic Soldiers.

With regard to the future use of the WVI, we suggest several steps be taken. First, we suggest that the WVI be administered experimentally in an operational selection context and a longitudinal validation effort be conducted. Although this chapter has clearly demonstrated the WVI has validity for predicting criteria in a concurrent sample, there are simply too many factors at play in an operational context (e.g., response distortion) which may attenuate the validity observed here to draw strong conclusions regarding how well the WVI would perform operationally. Indeed, previous Army research has demonstrated that the magnitude of differences between the psychometric properties of non-cognitive measures administered in operational and concurrent contexts can be substantial (Knapp, Waters, & Heggestad, 2002).

Second, as was apparent in the overview of the psychometric properties of the WVI, we have not provided reliability estimates for the scales that give rise to the WVI composites. Given the partially ipsative nature of the WVI, reporting internal consistency estimates for the WVI scales would be problematic (errors associated with the value pairs that comprise each scale score would be highly correlated). Ideally, future work will gather test-retest data on the WVI to assess (a) the consistency of individuals' preference for each reinforcer across occasions and (b) the consistency with which reinforcers are rank ordered by individuals across occasions.

A third consideration for future use of the WVI (which will be partially addressed in subsequent cross-instrument analysis chapters) is the potential benefit of combining information from the WVI and WPS to predict criteria of interest. Criterion-related validities for both the WVI composites summarized in this chapter and the WPS composites reported earlier were quite good, particularly for predicting the attitudinal criteria and Achievement and Effort performance. Furthermore, the correlations between the final WVI composites observed and the final WPS composites were only .46 (WVI: Unit ASat and WPS: Subjective AFit) and .34 (WVI: Unit AE and WPS: Unit AE). Taken together, these findings suggest that the WVI and WPS may be tapping enough unique variance such that when used in combination to predict the Select21 criteria, they have even more validity than evidenced in these chapters. Indeed, past research suggests that P-E fit measures that tap into multiple content domains (e.g., vocational interests, values, goals) exhibit higher levels of criterion-related validity than measures that focus on any single content domain (Kristof-Brown et al., 2005; O'Reilly, Chatman, & Caldwell, 1991).

A final consideration for future use of the WVI should be its potential utility for classification. Similar to our work on the WPS in developing a work values-based P-E fit measure for Select21, our primary focus was on assessing person-Army fit with regard to work values. This method runs contrary to how work values measures have traditionally been used in the vocational counseling and P-E fit literature. Typically, work value measures such as the MIQ and WIP (described earlier) have been used to assess fit to a particular occupation, vocation, or job (e.g., an MOS). We deviated from this tradition due to a generally held belief that the Army work environment provides a strong context that permeates the jobs of all first-term Soldiers, regardless of MOS. Similar to findings for the WPS in the previous chapter, the fact that the WVI was quite predictive of Army-wide criterion measures examined in this chapter (irrespective of MOS) suggests that this approach has merit. Nevertheless, these results should not be interpreted as meaning that measures of values-related "MOS fit" would fail to increment the validity of the values-related "Army fit" composites when predicting MOS-specific criteria. As such, we suggest future Army research, such as the research being conducted as part of ARI's Army Class project, assess whether WVI composites optimized within MOS offer any increment in validity over the more general person-Army fit composites described in this chapter.

CHAPTER 12: PSYCHOMOTOR TESTS

Teresa Russell, Huy Le, and Rod Rosse
HumRRO

Overview

Goal

The primary goal of the psychomotor tests was to increment the validity of the ASVAB for predicting certain aspects of future job performance in entry-level Army MOS and add classification efficiency to MOS assignment. Prior research supported the hypothesis that psychomotor measures would provide such incremental validity (McHenry & Rose, 1986). In several studies in the late 1980s, ARI found that the two tracking tests from Project A were useful predictors of gunnery performance (Grafton, Czarnolewski, & Smith, 1988; Smith & Graham, 1987; Smith & Walker, 1987). Later, research using the Project A Target Tracking tests (as a part of the Enhanced Computer-Administered Test [ECAT]) and ASVAB subtests (Sager, Peterson, Oppler, Rosse, & Walker, 1997) showed the usefulness of the tracking tests for enhancing classification efficiency. Using validation data from all the Services, the authors identified combinations of tests that were optimal for a specific purpose such as maximizing validity, minimizing adverse impact, and maximizing classification efficiency. These combinations of tests were called optimal batteries. Psychomotor tests from Project A appeared in all 20 of the optimal test batteries designed to maximize classification efficiency.

Two of the psychomotor tests from Project A were desirable for the Select21 project psychometrically and practically—Target Shoot and Target Tracking 1 (Russell et al., 2001). Both tests could be administered with one joystick; a full response pedestal like the one used in Project A was unnecessary.[48] In addition, both tests were designed to measure Psychomotor Precision (the ability to make muscular movements necessary to adjust or position a machine control mechanism) which subsumes Fleishman's (1967) Rate Control and Control Precision constructs.

Development Steps

Developing the initial versions of the psychomotor tests involved four steps: (a) selecting hardware, (b) developing test construction and delivery software, (c) pilot testing, and (d) field testing (which included a practice effects study). The specific procedures and results of the development work are described in detail in Russell, Katkowski, Le, and Rosse (2005). The most important findings from the development work were as follows:

- The psychomotor tests yielded highly reliable scores.
- The joysticks were comparable to each other in terms of the test scores they produced. The main effect of joystick was not significant in analyses.

[48] The customized Project A response pedestal had two joysticks and several buttons and dials. A picture appears in Campbell and Knapp (2001) on page 94.

- While there was a practice effect on these tests, it was not of great concern. The psychomotor test scores improved with practice, by one quarter to one-third of an *SD*, but improvements of that magnitude are often observed for cognitive tests (Russell, Reynolds, & Campbell, 1994). Additionally, the rank ordering of examinees' scores did not change much during the administration of a block of items, as indicated by reasonably high internal consistency estimates. The data suggested that the abilities that contribute to test performance are stable over the course of practice blocks. Specifically, relationships between test scores and ASVAB scores did not appear to change much with practice.

- The data and prior research suggested that it would be reasonable to combine the two Distance scores across the two tests to form a composite score (Psychomotor Precision) and retain the Time-to-Fire (latency) score as a separate score. The empirical rationale was that the two Distance Scores were correlated with each other (r = .51), and both improved with practice, while the Time-to-Fire score did not.

Instrument Description

Target Tracking Test

On each item of the Target Tracking test, a path consisting of vertical and horizontal line segments appears. A target box appears at the beginning of the path. A crosshair is centered in the box. As the item begins, the target starts to move along the path at a constant rate of speed. The examinee's task is to use a joystick to keep the crosshair centered within the target at all times. The concurrent validation version of this test has three practice items and nine scored items.

Target Shoot Test

At the beginning of an item on the Target Shoot test, a crosshair appears in the center of the screen and a target box appears at some other location on the screen. The target begins to move about the screen in an unpredictable manner, frequently changing direction. The examinee can control movement of the crosshair by using a joystick. The examinee's task is to move the crosshair into the center of the target and press a button on the joystick to "fire" at the target. The examinee must fire before the time limit on each trial is reached. This test has 3 practice items and 52 scored items.

Scoring

Description of Basic Scores

Target Tracking Test

The examinee's score on each item is a mean accuracy score—the average of the log distance from the center of the crosshair to the center of target taken every 50 milliseconds for the duration of the item. We constructed a total score on the test by computing the mean of the item Distance score means.

Target Shoot Test

The examinee receives three scores on each item. The first is a count of the number of "hits/misses/no fires" (i.e., a score indicating whether the examinee hit the target, missed the target or failed to fire at it). The second is a latency score—the time elapsed from the beginning of the trial until the examinee fires at the target. The third score is the distance from the center of the crosshair to the center of the target at the time the examinee fires at the target. Hits and Distance are both accuracy scores, with the Distance score being the more reliable of the two. Therefore, just two of the three basic scores were retained for subsequent analyses, Time-to-Fire (latency) and Distance (accuracy). Scores for the two retained scores were means across all items on the test.

Description of Final Scores

The psychomotor tests yielded two final scores. Psychomotor Precision is a composite created by adding the standardized Distance scores for the two tests. The Time-to-Fire (latency) score remains as a separate score. The empirical rationale was that the two Distance scores were correlated with each other (r = .51), and both improved with practice while the Time-to-Fire score did not. There was also support for this decision on the theoretical side. The two Distance scores were originally intended to tap Fleishman's (1967) two accuracy constructs, Rate Control and Control Precision. The Time-to-Fire score was added by the Project A team, and the team was not quite sure how this score fit in the psychomotor domain (Peterson, 1987). In exploratory factor analyses during Project A, it yielded split loadings on two factors, General Psychomotor and Perceptual Speed (which was defined by decision time scores on perceptual speed tests), with the loading on the General Psychomotor factor being slightly higher than the other loading.

Results

In the Select21 concurrent validation, 769 Soldiers took the psychomotor tests. Five cases were dropped due to incomplete or inappropriate data. Nine cases were dropped due to anomalies reported in the log that were severe enough to contaminate or distort the data (e.g., examinee wearing a cast on dominant arm, using other arm to respond). The final sample size was 755.

Psychometric Properties

Table 12.1 reports the means, standard deviations, and intercorrelations for the three basic scores and the Psychomotor Precision composite. Basic score statistics were comparable to those reported previously for these tests in the field test report (Russell et al., 2005).

Table 12.1. Descriptive Statistics for Psychomotor Basic Scores and Composite Scores

				Correlations		
Basic Score	*n*	*M*	*SD*	1	2	3
1. Target Tracking Distance Score (Mean Distance)	755	3.75	.53			
2. Target Shoot Distance Score (Mean Distance)	755	2.57	.27	**.62**		
3. Target Shoot Time-to-Fire (Seconds)	755	3.75	.95	**.34**	**-.10**	
4. Psychomotor Precision (sum of z-scores)	755	-.01	1.78	**.90**	**.90**	**.13**

Note. Statistically significant correlations are bolded ($p < .05$, two tailed).

Odd-even split half reliability estimates are reported in Table 12.2 for ease of comparison with reliabilities computed during Project A and prior Select21work. As shown, the Target Tracking Distance score was consistently highly reliable across several data collections, even when only a few items were administered. This finding was probably a result of the scoring process. During an item, the test measured the distance between the crosshairs and the target every 50 milliseconds. The score on an item was actually a mean of many Distance scores, making the overall Distance score a very reliable one. In contrast, the Distance score for the Target Shoot test was a point estimate (not a mean), and while its reliability was acceptably high, it was not as high as the reliability for the Target Tracking Test Distance score.

Table 12.2. Reliability Estimates for Psychomotor Test Scores

	Odd-Even Split-Half Reliability Estimates Corrected to Number of Items					Test-Retest Estimates		
	Select21			Project A[d]		Project A	Select21	Select21
	Concurrent Validation[a]	Field Test[b]	Pilot Test[c]	Incumbents (CV)	New Recruits (LV)	CV[d]	Pilot Test[c]	Field Test[b]
Target Tracking								
# Items	9	18	36	18	18	18	18	9
Distance Score	.96	.98	.97	.98	.98	.74	.87	.94[e]
Target Shoot								
# Items	54	60	60	30	30	30	30	30
Time-to-Fire	.93	.95	.92	.85	.84	.58	.81	.77
Distance Score	.89	.86	.85	.74	.73	.37	67	.64

[a] $n = 755$.
[b] $n = 637$. The tests were administered with no delay interval as a part of a practice effects study.
[c] $n = 119$. The tests were administered with no delay interval to study practice effects.
[d] *n*s for the Project A samples on which the split-half reliability estimates were computed were 9099-9274 (CV) and 6436 (LV). The *n* for the CV test-retest data was 473-479. The test-retest interval for the CV data was one month. LV data are from Peterson et al. (1992), and CV data are from Toquam et al. (1986).
[e] Adjusted to 18 items using the Spearman-Brown equation.

Criterion-Related Validity Estimates

We computed validity estimates for three scores, the two final scores (Time-to Fire and Psychomotor Precision), and the Distance score on Target Tracking. The two final scores were the best scores from the Select21 test battery. However, if the Army needs to shorten the battery significantly in the future, it may be desirable to use only the Target Tracking test, probably with more than nine items. We included the Distance score on Target Tracking to assess the possibility of using Target Tracking by itself.

We also reflected the psychomotor test scores. Recall that the psychomotor test scores are in latency and distance units. Low scores (faster, more accurate tracking) are better. For the validity analyses, we reflected all three scores by multiplying them by -1 in order to scale the scores in the more widely used direction (i.e., a high score is better). Validity computation and correction methods are described in detail in Chapter 6.

The raw and corrected zero-order validity estimates for predicting the Select21 performance and attitudinal criteria appear in Table 12.3. As might be expected, each of the three

psychomotor scores correlated highest with General Technical Proficiency. Psychomotor Precision and Target Tracking Accuracy, but not the Time-to-Fire score, also had significant correlations with Future Expected Performance. It is important to note that the Target Tracking Distance score had higher correlations with General Technical Proficiency and Achievement and Effort than the composite score (Psychomotor Precision) even though Target Tracking Distance only had nine items. The negative correlation between the Teamwork performance criterion score and the psychomotor test scores was unexpected. Since the Teamwork composite is the least reliable of the composites, the results may be spurious. There is no theoretical reason to expect that people who score high on the team construct would score low on the psychomotor ones.

Table 12.3. Criterion-Related Validity Estimates for Psychomotor Test Scores

	Performance Criteria					Attitudinal Criteria				
Predictor Scale	GTP	AE	PF	TEAM	FXP	ASat	AFit	CInt	ACog	FAA
	Uncorrected Validity Estimates									
Psychomotor Precision	**.17**	.03	.02	-.05	**.07**	.05	**.08**	-.02	**-.13**	**.14**
Target Shoot Time-to-Fire	**.10**	-.04	-.05	**-.09**	-.04	-.04	-.06	**-.09**	.03	-.03
Target Tracking Distance	**.19**	**.07**	.01	-.04	**.06**	.04	**.07**	-.03	**-.12**	**.11**
	Corrected Validity Estimates									
Psychomotor Precision	.25	.07	.02	-.06	.13	.05	.09	-.03	-.18	.13
Target Shoot Time-to-Fire	.20	.01	-.05	-.11	.02	-.04	-.06	-.11	-.01	-.04
Target Tracking Distance	.29	.12	.01	-.04	.14	.04	.08	-.04	-.18	.10

Note. n = 549 (AE criterion), n = 724 (all other performance criteria), n = 692-707 (attitudinal criteria). Corrected validity estimates have been corrected for criterion unreliability (first) and then indirect range restriction due to selection on the AFQT. Statistically significant correlations are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Unfortunately, we cannot directly compare these results to those from Project A. In Project A validity analyses were conducted at the predictor composite level, not at the individual test level. The Target Tracking and Target Shoot test scores were combined with scores from eight other tests in the analyses. While the computer test scores were very good predictors of General Soldiering Proficiency and Core Technical Proficiency (mean multiple correlations = .55 and.49 respectively), the validities across the computer battery do not tell us much about what we should expect for Select21.

Psychomotor test scores also yielded significant correlations with the weapons qualification score from the Personnel File Form (PFF). Uncorrected zero-order correlations between the psychomotor scores and the weapons qualification score were as follows: Psychomotor Precision (r = .13), Time-to-Fire (r = .12), and Target Tracking Accuracy (r = .13).

The psychomotor test also had some significant correlations with attitudinal criteria. For the Psychomotor Precision and Target Tracking Accuracy scores, better psychomotor performance was associated with better perceived fit, better attitudes about the future Army, and weaker intentions to leave the Army. This finding was surprising because the psychomotor test was not designed to predict attitudes. It is possible that its correlations with attitudinal measures

in the concurrent validation were indirectly related to the motivation of the participants. Another possibility is that a third variable such as gender affected the validity.

Incremental Validity Estimates

Incremental validity computation and correction methods are described in detail in Chapter 6, and the results for the psychomotor tests appear in Table 12.4. The psychomotor test contributed validity beyond the AFQT score for the prediction of General Technical Proficiency (.02 after corrections and adjustment for shrinkage). As discussed in the previous section, the incremental validity results for the Teamwork criterion may be spurious, and the significant incremental validities for some of the attitudinal criteria were unexpected.

Table 12.4. Incremental Validity Estimates for Psychomotor Test Scores

	Performance Criteria					Attitudinal Criteria				
Predictor Scale	GTP	AE	PF	TEAM	FXP	ASat	AFit	CInt	ACog	FAA
	Uncorrected Incremental Validity Estimates									
AFQT	**.30**	**.16**	.00	**.06**	**.17**	-.01	.00	**-.07**	**-.12**	-.05
Psychomotor Precision	**.03**	.00	.02	.02	.01	.04	**.07**	.00	**.05**	**.10**
Target Shoot Time-to-Fire	.00	.02	.05	**.06**	**.01**	.02	.06	**.03**	.01	.00
Target Tracking Distance	**.03**	.01	.01	.02	.00	.03	.07	.00	**.04**	**.08**
	Corrected Incremental Validity Estimates									
AFQT	.52	.28	.00	.16	.36	-.02	.01	-.11	-.23	-.08
Psychomotor Precision	.02	.00	.00	.01	.00	.00	.04	.00	.03	.08
Target Shoot Time-to-Fire	.00	.00	.00	.06	.01	.00	.00	.00	.00	.00
Target Tracking Distance	.02	.00	.00	.01	.00	.00	.02	.00	.02	.05

Note. n = 544 (AE criterion), n = 724-743 (all other performance criteria), n= 692-716 (attitudinal criteria). Cell values for the AFQT represent zero-order correlations between the AFQT and the given criterion (shown for reference). Uncorrected incremental estimates reflect the difference between the multiple R obtained when regressing the criterion on both the given composite and AFQT versus the R obtained when regressing the criterion only on the AFQT. Corrected incremental validity estimates reflect corrections for unreliability in the criterion (first), range restriction due to selection on the AFQT, and an adjustment for shrinkage using Rozeboom's (1978) formula. Statistically significant incremental validities are bolded ($p < .05$, one-tailed). GTP = General Technical Proficiency, AE = Achievement and Effort, PF = Physical Fitness, TEAM = Teamwork, FXP = Future Expected Performance, ASat = Satisfaction with the Army, AFit = Perceived Army Fit, CInt = Career Intentions, ACog = Attrition Cognitions, FAA = Future Army Affect.

Subgroup Differences

Tables 12.5 and 12.6 provide subgroup difference estimates (effect sizes) for gender and racial/ethnic comparisons, respectively. In both tables, the test scores were reflected such that higher scores indicated better performance. As shown, male Soldiers outperformed females by almost a full *SD* on accuracy/precision scores and by about 2/3 of an *SD* on the Time-to Fire score. These differences are consistent with those reported in Project A (Peterson et al., 1992).

Table 12.5. Psychomotor Test Scores by Gender

		Male		Female	
Psychomotor Score	d_{FM}	*M*	*SD*	*M*	*SD*
Psychomotor Precision Composite	**-0.95**	0.16	1.71	-1.46	1.76
Target Shoot Time-to-Fire	**-0.68**	-3.69	0.92	-4.32	1.06
Target Tracking Distance	**-0.95**	-3.70	0.52	-4.19	0.51

Note. n_{Male} = 683, n_{Female} = 71, d_{FM} = Effect size for Female-Male mean difference. Effect sizes calculated as (mean of females minus mean of males/*SD* of males. Statistically significant effect sizes are bolded, p <.05 (two-tailed).

As shown in Table 12.6, there was roughly one-third to two-thirds of an *SD* difference in psychomotor test scores between Black and White Soldiers. The standardized difference between Hispanic and White, Non-Hispanic mean scores was smaller and did not reach significance for the Psychomotor Precision score.

Table 12.6. Psychomotor Test Scores by Race/Ethnic Group

			White		Black		White Non-Hispanic		Hispanic	
Psychomotor Test Scores	d_{BW}	d_{HW}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Psychomotor Precision Composite	**-0.39**	-0.04	0.15	1.74	-0.53	1.83	0.16	1.75	0.10	1.71
Target Shoot Time-to-Fire	**-0.63**	**-0.40**	-3.63	0.91	-4.20	1.00	-3.56	0.90	-3.92	0.89
Target Tracking Distance	**-0.49**	**-0.20**	-3.69	0.52	-3.95	0.53	-3.67	0.51	-3.77	0.54

Note. n_{White} = 544, n_{Black} = 141, $n_{White\ Non\text{-}Hispanic}$ = 428, $n_{Hispanic}$ = 146, d_{BW} = Effect size for Black-White mean difference. d_{HW} = Effect size for Hispanic-White Non-Hispanic mean difference. Effect sizes calculated as (mean of minority group – mean of Whites) /*SD* of Whites. Statistically significant effect sizes are bolded, p < .05 (two-tailed).

Differential Prediction

Tables 12.7, 12.8, and 12.9 provide differential prediction results by gender, race, and ethnicity, respectively. In all of the analyses, psychomotor scores were reflected so that higher scores indicate better performance.

As shown by the bolded values in Table 12.7, 11 of the 30 intercept tests and one of 30 slope tests were significant for gender. There were no significant differences between male and female slopes and intercepts for General Technical Proficiency, the criterion most related to psychomotor test scores. Where differences were found they indicated that the psychomotor test scores tend to underpredict females' Achievement and Effort, Teamwork, and Future Expected Performance scores and Attrition Cognitions.

As shown by the bolded values in Table 12.8, 13 of the 30 intercept tests and none of 30 slope tests were significant for race. The psychomotor test scores tended to overpredict Black Soldiers' Achievement and Effort, Teamwork, and Future Expected Performance scores. They tended to underpredict Attrition Cognitions for Black Soldiers.

As shown by the bolded values in Table 12.9, 6 of the 30 intercept tests and none of the 30 slope tests were significant for ethnicity. The significant intercepts suggest that psychomotor test scores underpredict Hispanic Soldiers' Teamwork and Future Army Affect scores.

Table 12.7. Differential Prediction Results for Psychomotor Scores by Gender

	Psychomotor Precision Composite					Target Shoot Latency					Target Tracking Distance				
	Gender	Precision *b*		*r* by Gender		Gender	Latency *b*		*r* by Gender		Gender	Distance *b*		*r* by Gender	
	b	M	F	M	F	*b*	M	F	M	F	*b*	M	F	M	F
General Technical Proficiency	.13	.09	.16	**.16**	**.31**	.03	.05	.06	**.10**	.13	.16	.09	.20	**.18**	**.36**
Achievement and Effort	**.30**	.05	.04	.09	.09	**.24**	.00	-.04	.00	-.08	**.33**	.06	.08	**.12**	.15
Physical Fitness	-.06	.00	.06	.01	.08	-.15	-.04	-.06	-.06	-.09	-.11	.00	.00	.00	.00
Teamwork	**.20**	-.02	.02	-.03	.03	.12	-.03	-.12	-.05	-.20	**.21**	-.01	.03	-.02	.04
Future Expected Performance	**.29**	.05	.12	**.08**	.19	.18	-.02	-.01	-.02	-.01	**.31**	.05	.15	**.08**	.22
Satisfaction with the Army	-.18	.02	.05	.03	.07	**-.28**	-.04	-.08	-.05	-.13	-.20	.01	.04	.01	.05
Perceived Army Fit	.19	**.05**	**.26**	.05	**.30**	-.17	-.03	-.21	-.03	**-.27**	.11	.05	.16	.06	.18
Attrition Cognitions	.23	-.11	-.11	**-.11**	-.11	**.49**	.03	.24	.03	**.26**	**.32**	-.11	.00	**-.11**	.00
Career Intentions	.01	-.02	.02	-.02	.02	-.23	-.07	-.34	-.06	**-.33**	-.11	-.02	-.12	-.02	-.10
Future Army Affect	-.10	.10	.24	**.10**	**.26**	**-.42**	-.02	-.19	-.03	-.24	-.22	.08	.09	**.09**	.09

Note. $n_{Regression}$ = 548-723; n_{Male} = 488-659; n_{Female} = 60-69. Gender *b* = Unstandardized regression weight for gender (0 = male, 1 = female). Psychomotor score *b* = Unstandardized regression weight for the given psychomotor score for males and females. *r* by Gender = Correlation between the given psychomotor score and the given criterion for each gender. Regression weights for males and females are bolded if the score-by-gender interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for gender are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

.

Table 12.8. Differential Prediction Results for Psychomotor Scores by Race

	Psychomotor Precision Composite					Target Shoot Latency					Target Tracking Distance				
	Race	Precision *b*		*r* by Race		Race	Latency *b*		*r* by Race		Race	Distance *b*		*r* by Race	
Criterion	*b*	W	B	W	B	*b*	W	B	W	B	*b*	W	B	W	B
General Technical Proficiency	**-.23**	.10	.03	**.18**	.08	**-.25**	.04	-.02	.08	-.05	**-.22**	.09	.06	**.17**	.12
Achievement and Effort	**-.15**	.04	-.06	.07	-.11	**-.17**	-.03	-.06	-.05	-.12	**-.14**	.04	-.02	.09	-.05
Physical Fitness	.01	.02	.03	.03	.04	-.04	-.04	-.08	-.05	-.11	.00	.01	.00	.01	-.01
Teamwork	-.02	-.01	-.06	-.01	-.10	-.03	-.06	-.04	**-.10**	-.06	-.02	-.01	-.03	-.02	-.06
Future Expected Performance	**-.13**	.06	.02	.08	.04	**-.17**	-.03	-.05	-.05	-.09	**-.13**	.05	.02	.07	.03
Satisfaction with the Army	-.03	.05	.04	.06	.05	-.05	-.07	.02	**-.09**	.02	-.01	.02	.07	.03	.08
Perceived Army Fit	-.10	.06	.09	.07	.12	-.17	-.10	-.04	**-.12**	-.06	-.09	.03	.10	.04	.12
Attrition Cognitions	**.35**	-.13	-.02	**-.12**	-.03	**.43**	.06	.10	.06	.11	**.35**	-.10	-.03	**-.10**	-.03
Career Intentions	.04	.01	-.03	.01	-.03	-.01	-.11	-.09	**-.10**	-.09	.01	.00	-.12	.00	-.11
Future Army Affect	-.10	.10	.23	**.11**	**.25**	**-.22**	-.07	-.04	-.07	-.04	-.10	.06	.20	.07	**.20**

Note. $n_{Regression}$ = 501-656; n_{White} = 396-524; n_{Black} = 105-132. Race *b* = Unstandardized regression weight for race (0 = White, 1 = Black). Psychomotor score *b* = Unstandardized regression weight for the given psychomotor score for Whites and Blacks. *r* by Race = Correlation between the given psychomotor score and the given criterion for each race. Regression weights for Whites and Blacks are bolded if the score-by-race interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for race are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Table 12.9. Differential Prediction Results for Psychomotor Scores by Ethnic Group

	Psychomotor Precision Composite					Target Shoot Latency					Target Tracking Distance				
		Precision *b*		*r* by Ethnicity			Latency *b*		*r* by Ethnicity			Distance *b*		*r* by Ethnicity	
	Ethnicity					Ethnicity					Ethnicity				
Criterion	*b*	W	H	W	H	*b*	W	H	W	H	*b*	W	H	W	H
General Technical Proficiency	-.03	.10	.02	**.19**	.05	-.02	.03	.08	.06	.15	-.03	.10	.02	**.18**	.05
Achievement and Effort	.07	.04	.00	.08	-.01	.06	-.04	.04	-.07	.08	.07	.04	.04	.08	.08
Physical Fitness	.11	.02	.02	.02	.03	.09	-.05	-.02	-.06	-.03	.11	.00	.03	.00	.04
Teamwork	**.14**	-.03	.03	-.05	.06	**.13**	-.07	.01	**-.11**	.01	**.14**	-.03	.04	-.06	.07
Future Expected Performance	.06	.06	.01	.08	.01	.05	-.05	.05	-.07	.08	.06	.04	.02	.06	.04
Satisfaction with the Army	.14	.07	-.03	.09	-.04	.12	-.08	.01	-.09	.02	.14	.04	-.01	.05	-.01
Perceived Army Fit	.12	.06	.02	.07	.03	.10	-.12	.04	**-.13**	.04	.12	.03	.04	.04	.05
Attrition Cognitions	-.02	-.14	-.08	**-.14**	-.08	.01	.07	.00	.07	.00	-.03	-.11	-.08	**-.11**	-.09
Career Intentions	.04	.00	-.02	.00	-.02	-.01	-.13	-.12	**-.11**	-.11	.04	.00	-.01	.00	-.01
Future Army Affect	**.23**	.10	.12	.09	.14	**.21**	-.07	.01	-.07	.01	**.24**	.08	.06	.08	.08

Note. $n_{Regression}$ = 414-552; $n_{White,non-Hispanic}$ = 313-411; $n_{Hispani\ c}$ = 101-141. Ethnicity *b* = Unstandardized regression weight for ethnicity (0 = White non-Hispanic, 1 = Hispanic). Psychomotor score *b* = Unstandardized regression weight for the given psychomotor score for White non-Hispanics and Hispanics. *r* by Ethnicity = Correlation between the given psychomotor score and the given criterion for each ethnic group. Regression weights for White non-Hispanics and Hispanics are bolded if the score-by-ethnicity interaction is statistically significant ($p < .05$, two-tailed). Statistically significant regression weights for ethnicity are bolded ($p < .05$, two-tailed). Statistically significant correlations are bolded ($p < .05$, one-tailed).

Summary

Key Findings

Validity

A fairly strong body of evidence has accumulated for the validity of psychomotor tests for predicting some criteria. In several studies in the late 1980s, ARI found that the two tracking tests from Project A were useful predictors of gunnery performance (Grafton et al., 1988; Smith & Graham, 1987; Smith & Walker, 1987). Later, research using the Project A Target Tracking tests (as a part of the Enhanced Computer-Administered Test [ECAT]) and ASVAB subtests (Sager et al., 1997) showed the usefulness of the tracking tests for enhancing classification efficiency. Using validation data from all the Services, the authors identified combinations of tests that were optimal for a specific purpose such as maximizing validity, minimizing adverse impact, and maximizing classification efficiency. These combinations of tests were called optimal batteries. Psychomotor tests from Project A appeared in all 20 of the optimal test batteries designed to maximize classification efficiency. Select21 CV validation results were consistent with prior research. Specifically, each of the three psychomotor scores correlated highest with General Technical Proficiency and provided incremental validity for predicting that criterion.

Subgroup Differences and Differential Prediction

Consistent with previous findings on these tests, male-female subgroup differences were large—nearly one *SD* difference with males receiving the higher scores. Race and ethnic group differences were typically one-half *SD* or less, with White Soldiers receiving the higher scores. Across all of the predictive bias analyses, differences were primarily in intercepts (i.e., 30 of 90 intercept tests were significant while only one slope test was significant). Intercept differences usually indicated underprediction of female Soldiers' performance scores and overprediction of Black Soldiers' performance scores.

Issues Regarding Operational Use

Validation and Classification Efficiency

Based on our results and those from the ECAT and Project A projects, we expect the psychomotor tests to add classification efficiency (i.e., increase mean predicted performance) for some MOS. In addition to predicting gunnery performance, research suggests that psychomotor skills are important for operating uninhabited combat aerial vehicles (Kay, Dolgin, Wasel, Langelier, & Hoffman, 1999) and, of course, aviators (North & Griffin, 1977; Street & Dolgin, 1994). Additional research needs to estimate classification gains for the entry-level MOS that require psychomotor abilities.

Administration Time

Administration time is always an important consideration in experimental and operational testing. Tests are more palatable if they do not require a lot of examination time. Both tests are

self-paced; together they take about 21 minutes on average of administration time (i.e., 4 min. for Tracking instructions, 4 min. for 9 Tracking items, 4 min. for Target Shoot instructions, and 9 min for 52 Target Shoot items). Clearly, a Target Shoot item takes much less time than a Tracking item does to administer. However, one of the more important findings was that the Target Tracking Distance score had high levels of validity and useful incremental validity by itself. Even with 52 items, the Target Shoot Test scores did not achieve the high reliability of the nine item Target Tracking Test Distance score. Also, the Target Tracking Distance Score was at least as valid as the composite and the other scores under consideration. The administration time for the tests could be reduced by eliminating the Target Shoot Test. If administration time permits, adding items to the Target Tracking test (to mitigate potential practice effects) is likely to be a better use of administration time.

Response Apparatus

The problem that the Army has had in trying to implement psychomotor tests has to do with the apparatus. In Project A, the response pedestal was designed and produced to meet specifications. But, it was fairly large, difficult to transport, and required periodic repairs. In the Select21 project, we attempted to simplify the apparatus and associated workload by using modified commercial, off-the-shelf joysticks. It is likely that the Project A response pedestals were more durable than the commercial joysticks; several of our joysticks had become unusable by the end of the Select21 validation data collection. But, all-in-all, our efforts were successful. Using the commercial joysticks, we obtained estimated validities and reliabilities that were comparable to those from Project A.

Even though the use of commercial joysticks was reasonably successful, we expect that the major obstacle to implementing the psychomotor tests will continue to be related to the purchase and maintenance of an apparatus. A commonplace, multipurpose apparatus, such as a mouse, would be easiest to implement because it is already standard equipment. Therefore, we recommend additional research to assess the construct validity of scores on tracking tests with internal or external mice. This research should have a within-subjects design such that all subjects would take a joystick version of the test and a mouse version. The order of administration (i.e., joystick or mouse) should be counterbalanced across subjects. The results of the study would indicate whether a mouse can be used to reliably measure psychomotor precision.

CHAPTER 13: CROSS-INSTRUMENT ANALYSES AND RESULTS

Jennifer L. Burnfield, Teresa L. Russell, and Dan J. Putka
HumRRO

Overview

In this chapter, we examine relations among the predictor measures. These cross-instrument analyses had three main goals. The first was to provide construct validity evidence (beyond that provided in the individual instrument-specific chapters) for the scales comprising the predictor instruments. The second goal was to identify areas of redundancy and uniqueness, particularly among measures of similar constructs or domains. Such analyses can inform practical considerations and recommendations for operational use. The third objective was to examine the incremental validity that each Select21 predictor measure offers over the Armed Forces Qualification Test (AFQT), as well as over the combination of AFQT, ASVAB Technical Composite (see Chapter 6), and ASVAB Spatial. The purpose of the latter incremental validity analyses is to assess the validity increment of Select21 predictors over not only the current selection composite (AFQT), but also other potentially viable ASVAB based selection measures.

Construct Validity Evidence

This section discusses relationships among the Select21 predictor scales in an attempt to expand the nomological network for the constructs they measure. To this end, we organized the predictors into five conceptually-driven individual differences domains:

- Cognitive Ability—includes Armed Services Vocational Aptitude Battery (ASVAB) scores.
- Psychomotor Ability—includes the Target Tracking (TT) Distance score.
- Judgment—includes the Predictor Situational Judgment Test (PSJT) judgment score.
- Temperament—includes the Work Suitability Inventory (WSI) and Rational Biodata Inventory (RBI).
- Interests/Work Values—includes the Work Preferences Survey (WPS) and the Work Values Inventory (WVI).

We focus specifically on relationships among predictor variables that have theoretical importance for particular constructs. We also report only results that have been corrected for range restriction, because they are better estimates (than raw correlations) of the population-level relationships among constructs. Full tables of raw and corrected-for-range restriction correlations appear in Appendix C.

Cognitive Ability

There is a relatively large body of literature focused on the construct validity of the ASVAB. In short, the three most important findings are: (a) hierarchical factor analyses have found that the general factor (psychometric *g*) accounts for approximately 60% of the total variance (Kass, Mitchell, Grafton, & Wing, 1983; Welsh, Watson, & Ree, 1990), (b) non-hierarchical factor analyses have identified four factors which have been replicated across studies

(Kass et al., 1983; Welsh, Kucinkas et al., 1990), and (c) the four factors have been replicated for males, females, Blacks, Whites, and Hispanic subgroups separately (Kass et al., 1983). The four factors and the ASVAB subtests that have substantial loadings on them are:

(1) Verbal (Word Knowledge [WK] and Paragraph Comprehension [PC])
(2) Speed (Coding Speed [CS] and Numerical Operations [NO])[49]
(3) Quantitative (Arithmetic Reasoning [AR] and Math Knowledge [MK])
(4) Technical (Auto and Shop [AS], Mechanical Comprehension [MC], and Electronics Information [EI])

In non-hierarchical factor analyses, the General Science (GS) subtest tends to load on the Verbal factor (Ree, Mullins, Mathews, & Massey, 1982), and has yielded split-loadings on the Verbal and Technical factors (Kass et al., 1983). With the exception of GS results, this factor solution is relatively straightforward and is highly replicable. Even so, over half of the variance in ASVAB scores is accounted for by the general factor (Welsh, Watson et al., 1990).

Given the relatively large body of literature on the ASVAB's construct validity, we focus on the Spatial test (Assembling Objects) in the following paragraphs. It is a new addition to the ASVAB and is used experimentally by the Army at this time. Below, we provide some historical context for AO and discuss its relationships with other cognitive measures. Temperament, interest, and values constructs do not have a strong theoretical link to spatial ability and are not therefore discussed here.

Six spatial tests (one of which was AO) were developed during Project A (Campbell & Knapp, 2001). Project A data suggested that AO was a good candidate for inclusion in the ASVAB because it (a) had high loadings on a general spatial factor when the spatial test correlations were hierarchically factored (that is, it appear to be a broad and general spatial test) and (b) had relatively low subgroup differences compared to the other spatial tests (Peterson et al., 1992). Russell, Reynolds, and Campbell (1994) present a summary of results for AO accumulated across the Army's Project A and the joint-service Enhanced Computer Adaptive Test (ECAT) project.

Historically, the study of spatial ability is linked to the study of mechanical aptitude, and studies typically report a high correlation between indicators of the two constructs (Bennett, Seashore, & Wesman, 1974; Guilford & Lacey, 1947). In the Select21 concurrent validation (CV) sample, the Spatial score correlated $r = .58$ (corrected) with the Technical composite. Spatial was also correlated with AFQT, $r = .55$ (corrected). The advantage offered by spatial ability tests is that they can measure constructs that are conceptually related to mechanical ability while yielding lower gender differences than mechanical tests often do. As noted in Chapter 6, the gender difference on Spatial was substantially smaller than the gender difference on the Technical composite. Chapter 6 also provided evidence that the Spatial score added validity beyond that provided by the AFQT together with the Technical composite.

[49] Coding Speed and Numerical Operations are currently used only by the Navy.

Psychomotor Ability

As mentioned in Chapter 12, the Target Tracking (TT) Distance score was designed to be a measure of psychomotor precision (Campbell & Knapp, 2001; Russell, Katkowski, Le, & Rosse, 2005)—the ability to make muscular movements necessary to adjust or position a machine control mechanism—which subsumes Fleishman's (1967) Rate Control and Control Precision constructs. The Target Tracking test was developed during the Army's Project A and was also administered as a part of the joint-Service ECAT study (Russell et al., 1994). Theoretically, psychomotor ability involves spatial processing of the stimulus and motor control; it is linked to spatial ability and motor abilities. TT Distance's relationships with temperament, interest, and values variables are not focal to the psychomotor construct and are not discussed here.

Historically, research indicates that psychomotor abilities are related to cognitive ability, particularly those within the spatial domain (Fleishman, 1967), and technical skills such as mechanical comprehension (Ree & Carretta, 1995). For example, the psychomotor test scores were significantly correlated with spatial test scores (r = .48), ASVAB Technical scores (r = .55), and ASVAB Quantitative scores (r = .58) in a sample of more than 4,000 Soldiers in Project A (Peterson et al., 1992). Similarly, in the Select21 CV, the TT Distance score correlated more highly with Spatial and Technical than with AFQT (i.e., r = .36 [corrected] with Spatial, r = .40 [corrected] with Technical and r = .29 [corrected] with AFQT).

Psychomotor precision involves motor skills or muscular movements (Fleishman, 1967; McHenry & Rose, 1986). Psychomotor abilities appear to have a small positive relationship with self-reported physical fitness/motivation but little relationship to actual physical fitness measures. Psychomotor scores yielded significant but small correlations with self-reported fitness scales including the RBI Fitness Motivation scale (Select21) and the Physical Condition scale from the Assessment of Background and Life Experiences (ABLE; Peterson et al., 1992). Even so, psychomotor precision was not related to measures of physical fitness in the Select21 sample.[50]

Judgment

The debate concerning what constructs situational judgment tests (SJTs) measure has been ongoing for over 75 years (see Moss & Hunt, 1926; Schmitt & Chan, 2006; Thorndike, 1936). One point of agreement is that SJTs reflect a measurement method, and that choices made by test developers drive the specific content focus of any given SJT (Weekley & Ployhart, 2006). Schmitt and Chan (2006) have suggested that at the highest level, SJTs simply measure judgment. Research suggests that virtually all SJTs have a strong interpersonal component and some relationship with cognitive ability (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). The PSJT Judgment score's relationships with cognitive and temperament constructs are described below.

[50] Body Mass Index (BMI) is an indicator of body fatness which is used as a surrogate measure of fitness (http://www.cdc.gov/nccdphp/dnpa/bmi/), and the Army Physical Fitness Test (APFT) is an indicator of fitness. In the Select21 CV sample, BMI and APFT scores were correlated significantly in the predicted direction (r = -.21 p < .05, two tailed). Regardless, neither BMI nor APFT had a significant relationship to the Target Tracking Distance score.

The PSJT Judgment score appears to be slightly less correlated with cognitive ability than other SJTs. A recent meta-analysis (McDaniel et al., 2001) reported a mean observed correlation between SJT scores and cognitive ability of .36 (corrected to .46), but there was quite a bit of variance in the correlations. The PSJT Judgment score correlated .34 (corrected) with AFQT, .20 (corrected) with ASVAB Spatial, and .21 (corrected) with ASVAB Technical. Notably, the PSJT Judgment score correlated more strongly with the AFQT, which measures verbal and mathematical abilities, than with the ASVAB scores measuring other abilities.

The PSJT results were consistent with prior research suggesting that SJTs (a) often yield moderate correlations with temperament measures, and (b) tend to show their strongest relationships with three of the Big Five constructs—Agreeableness, Emotional Stability, and Conscientiousness (McDaniel & Nguyen, 2001; Schmitt & Chan, 2006). The PSJT Judgment score was most highly correlated with the following two scales from the RBI: Hostility to Authority (r = -.38 corrected) and Cultural Tolerance and Gratitude Toward Others (r = .32 corrected). These RBI scales are strongly related to Agreeableness (Kilcullen, Putka, McCloy, & Van Iddekinge, 2005). Judgment scores were also positively correlated with the RBI Internal Locus of Control, Cognitive Flexibility, and Achievement Orientation scores. Earlier research has noted (Kilcullen et al., 2005) that RBI Internal Locus of Control is related to Emotional Stability and that Achievement Orientation is related to Conscientiousness. While several of the RBI scales correlated significantly with the PSJT Judgment score, only two of 16 correlations between Judgment and WSI scales were significant (r = -.11 with Stress Tolerance and r = -.09 with Persistence).

According to Schmitt and Chan (2006), the correlations between SJT scores and interest measures are likely to be a function of SJT content. For example, they report some evidence that scores on SJTs that consist of knowledge- or learning-oriented content are related to Investigative interests, and SJTs with interpersonal content are related to Social interests. The PSJT Judgment score yielded small but significant correlations with Social, Conventional, Investigative and Enterprising interests, the highest of which were with Social interests (r = .18 corrected). It was not correlated with Realistic or Artistic interests from the WPS. Since most of the situations on the PSJT involve social or team instructions, the higher correlation with Social interests makes sense.

The PSJT Judgment score appears to be strongly related to work values, as it correlated positively and significantly with 26 of the 28 WVI scales. The highest correlations (corrected r = .20 or greater) were with Social Service, Ability Utilization, Emotional Development, Societal Contribution, Leadership Opportunities, Advancement, Esteem, Autonomy, Co-Workers, Personal Development, Home, and Achievement.

While the PSJT is correlated with cognitive ability, its correlations with temperament constructs, values, and interests suggest that it is measuring more than *g*. Based on its correlations with the RBI and other scales, it appears to tap achievement motivation and perhaps interpersonal skills.

Temperament

The RBI and the WSI were the two Select21 predictors designed to tap temperament-oriented constructs. Even so, it is important to note that the RBI and the WSI represent two very different measurement methods.[51] In this section, we report WSI scale scores (as opposed to the composite scores) which are completely ipsative (i.e., the sum of full WSI scores on each dimension will be a constant for all respondents). Given the nature of ipsative data, it is important to note that a respondent's scores on the WSI scales do not reflect normative standing on the trait of interest (i.e., the respondent's level of Attention to Detail *relative to other Soldiers* in their sample). Rather, the WSI scores reflect a respondent's judgment regarding his or her ability to perform the type of work described by a given WSI statement *relative to the types of work* described by the other WSI statements (i.e,. how well the respondent thinks he or she would perform types of work requiring Attention to Detail relative to types of work requiring other traits). Thus, correlations between WSI scales and other Select21 measures index the extent to which the other Select21 measures are related to Soldiers' perceived relative strengths and weaknesses when it comes to the non-cognitive demands of Army work. Therefore, the nature of the correlations among WSI scale scores, and between the WSI scale scores and scores on other measures, may occasionally seem counterintuitive.

Some RBI and WSI scale score correlations were consistent with our expectations; others were not. Logically, WSI Cultural Tolerance was moderately related to RBI Cultural Tolerance (r = .34 corrected). Similarly, it makes sense that WSI Achievement/Effort was related to RBI Achievement (r = .12 corrected), RBI Self Efficacy (r = .13 corrected), and RBI Internal Locus of Control (r = .12 corrected). In addition, those who scored higher on WSI Concern for Others tended to score lower on RBI Narcissism (r = -.10 corrected). However, it is unclear why WSI Concern for Others had moderate negative relations with most of the other RBI scores (e.g., with Army Identification, Self Efficacy, and Fitness Motivation: corrected rs = -.26, -24, and -.29, respectively).

As might be expected, WSI Independence scores (high scorers indicated they would be more effective at types of work that required independence) showed several negative relationships with RBI scales that pertain to interacting with others (e.g., corrected rs = -.17, -.14, and -.13 for Interpersonal Skills-Diplomacy, Cultural Tolerance, and Respect for Authority, respectively). WSI Leadership Orientation was positively related to a number of RBI scales, most strongly with Peer Leadership as might be expected (corrected r = .27). Contrary to expectations, WSI Initiative was not related to RBI Achievement or Self-Efficacy. Again, this may be due to the ipsativity of the WSI item-level scores.

Even though correlations between temperament scales and cognitive ability are often small but significant (see Project A correlations; Peterson et al., 1992), specific personality and biodata scales do correlate with cognitive ability in meaningful ways. For example, Judge, Higgins, Thoresen, and Barrick (1999) reported significant raw correlations between general mental ability and Openness to Experience (r = .33 [n = 194 adults] and .41 [n = 166 children]),

[51] The RBI is a biodata inventory with rationally developed scales, while the WSI asks respondents to rank order cards (relating to traits) in terms of "how well you think you would perform the type of work described by the cards."

Conscientiousness (r = .29 [n = 194 adults] and .53 [n = 166 children]), and Neuroticism (r = -.22 [n = 194 adults] and -.43 [n = 166 children]). Similarly, the highest correlation between any of the RBI scales and AFQT in the Select21 sample was with the RBI Cognitive Flexibility scale (corrected r = .47), which is related to Openness to Experience (Kilcullen et al., 2005). The RBI's Peer Leadership, Stress Tolerance, Self-Efficacy, Internal Locus of Control, and Gratitude Toward Others scales also correlated significantly and positively with AFQT, with corrected correlations in the .18 to .24 range. People who received higher Lie Scale and Hostility to Authority scores on the RBI tended to have lower AFQT scores (corrected rs = -.19 and -.24, respectively). For the WSI, Independence, Innovation, and Stress Tolerance yielded significant positive correlations with AFQT (corrected rs = .25, .18, and .15, respectively) as might be expected. Cooperation, Concern for Others, and Energy yielded significant negative correlations with AFQT (corrected rs = -.26, -.23, and -.13).

Correlations between the WSI full scale scores and the interest and values measures demonstrated some evidence of convergent validity. WSI Independence was positively related to WVI Independence and inversely related to the WPS Social Interests scale and facet scores (Working with Others, Helping Others). In addition, the WSI Innovation score was positively associated with the WPS Artistic Interests scale and facet scores (Artistic Activities, Creativity; r = .23-.32) and to WVI Creativity (r = .22). Similarly, WSI Attention to Detail was associated with the WPS Conventional Interests scale and facet scores, the strongest relationship being with the WPS Detail Orientation facet. WSI Energy was moderately correlated (r = .36) with the WPS Physical facet and the WVI Physical Development scale (r = .26). The WSI Leadership Orientation scale was positively related to the WPS Lead Others facet and WVI Leadership Opportunities, while showing weaker or non-significant relationships with the other variables.

The RBI tended to show stronger relationships with interests than with values. Many of the RBI scale scores had significant, positive, moderate-to-strong correlations with the WPS scale and facet scores and had less consistent and weaker relationships with the WVI scale scores. The strongest relationships were observed between RBI Cognitive Flexibility and the WPS Investigative Interests scale (r = .55 [corrected]) and Critical Thinking facets (r = .58 [corrected]), Conduct Research facet (r = .40 [corrected]) and the WPS Creativity facet (r = .45 [corrected]). This provides some evidence of convergent validity, as people high on Cognitive Flexibility may be drawn to the creative problem solving tasks that are inherent in investigative activities. In addition, RBI Achievement was moderately correlated with the WPS Social Interest scale and facets (with corrected rs ranging from .17 to .30). Perhaps Working with Others and Helping Others are ways of attaining goals in the Army given the team-oriented nature of Army work (e.g., combat units). Moderate relationships were also observed between RBI Interpersonal Skills-Diplomacy and the WPS Social Interests scale and Work with Others facet (corrected rs = .41 and .44, respectively).

Overall, the counterintuitive results presented in this section appeared most frequently with respect to the RBI-WSI relationship and were less notable for values and interest scores. In part, this is due to the fact that we had a number of specific expectations for the WSI-RBI relationships. It is, however, very important to note that McCloy and Putka (Chapter 8) have recommended use of the WSI's empirical dyad scoring (which reduces the effect of ipsativity). They noted that the empirical keying approach showed promise, but further research is needed to support the WSI validation and scoring work as described in Chapter 8.

Interests and Work Values

The Select21 P-E fit predictor measures were designed to assess respondents' interests (WPS) and work values (WVI). A number of conceptually plausible inter-relations among these measures' scales were observed. For instance, the WVI Physical Development scale was moderately associated with the WPS Realistic scale (corrected r = .30) and Physical facet (corrected r = .40). In addition, the WVI Creativity scale was positively associated with the WPS Artistic Interests scale (corrected r = .20), as well as with the Artistic Activities and Creativity facets (corrected rs = .14 and .23, respectively). Not surprisingly, the WPS Social Interests scale was positively associated with the WVI Societal Contribution scale (corrected r = .24) and with the WVI Social Service scale (corrected r = .32) but negatively associated with WVI Independence (corrected r = -.27). Finally, the WPS Enterprising Interests scale and the Prestige and Leading Others facets were positively associated with WVI scales that also purport to assess leadership (Leadership Opportunities), status (Social Status), and career progression (Advancement).

Correlations among cognitive abilities and the interest and value measures appeared to be meaningful. For example, WPS Realistic Interests were negatively correlated with AFQT (corrected r = -.21) and positively correlated with ASVAB Technical (corrected r = .10). Similarly, Project A's Rugged/Outdoors interest scale (i.e., a realistic interest scale) correlated .36 with ASVAB Technical while correlating only .18 with an ASVAB verbal composite (Peterson et al., 1992). Several of the WVI scales correlated significantly with AFQT; the highest two correlations were with Ability Utilization (corrected r = .32) and Autonomy (corrected r = .29).

A number of the interest scores were significantly associated with PSJT Judgment, the psychomotor scores, and the temperament scales in ways that are consistent with expectations for the interest constructs. For example, PSJT Judgment correlated highest with Social interests (corrected r = .18). Many of these relationships were discussed in previous sections of the chapter, so they are not reiterated in detail here. The strongest relationships were between the WPS scale scores and the RBI (e.g., WPS Investigative Interests scale and facets with RBI Cognitive Flexibility). Relationships were generally weaker between the WPS and WSI scale scores.

Regarding the WVI correlations with other predictors, the associations were generally modest. As with the WPS, there were relatively few significant associations between the WVI and WSI. The strongest relationships were between WVI Leadership and WSI Leadership Orientation (corrected r = .28), between WVI Leadership and RBI Achievement (corrected r = .26), and between WVI Physical Development and WSI Energy (corrected r = .26).

Scale Correlations Summary

In general, the correlations between scales were consistent with prior research that had employed similar measures or constructs. In particular, the spatial, psychomotor, interest and values measures showed expected patterns of correlations with scores from other domains. An exception was the ipsatively-scored WSI scales, which did not yield the expected pattern of inter-correlations, especially with relevant RBI scale scores. However, as noted previously, the WSI scale scores were not designed to measure an individual's normative standing on a trait.

Composite Correlations

The previous section discussed correlations among predictor *scales* for the purpose of enhancing understanding of the constructs measured by the instruments. The purpose of this section is to assess the uniqueness/redundancy of the instruments by comparing correlations among predictor *composites*. The focus is primarily on potential redundancy among the WVI, WPS, and WSI composites.

WVI, WPS, and WSI Composite Formation

Composite scores were computed for three Select21 measures: the WVI, WPS, and WSI. Complete descriptions of how these composites were formed are presented in previous chapters of this report, but are revisited here for convenience. Ten composite scores were computed for the WSI. As discussed in Chapter 8, these WSI composites comprised dyad-level scores (i.e., dummy variables indicating whether a given WSI dimension was ranked higher than another dimension) that were selected to optimally predict the 10 criterion composite scores (see Table 8.8 for a description of the dyads that contribute to each composite).

As summarized in Chapter 10, the WPS had two final composite scores: a unit-weighted composite of WPS facets targeting Achievement and Effort (WPS Unit AE) and a subjectively-weighted composite of WPS facets targeting Perceived Fit with the Army (WPS Subjective AFit). The WPS Unit AE composite consisted of scores from the Critical Thinking, Artistic Activities, Help Others, and Detail Orientation facets. The WPS Subjective AFit composite was comprised of weighted scores from the Physical, Creativity, Help Others, Work with Others, High Profile, Lead Others, and Clear Procedures facets.

Similar to the WPS, two final composites scores were calculated for the WVI: a unit-weighted composite of WVI scales targeting Achievement and Effort (WVI Unit AE), and a unit-weighted composite of WVI scales targeting Satisfaction with the Army (WVI Unit ASat). As described in Chapter 11, the WVI Unit AE composite consisted of scores from the Emotional Development, Independence, Leadership Opportunities, Leisure Time, and Societal Contribution scales. The WVI Unit ASat composite comprised scores from the Ability Utilization, Advancement, Comfort, Creativity, Emotional Development, Flexible Schedule, Independence, Leisure Time, Physical Development, Social Status, and Travel scales.

Raw and corrected correlations were computed among the composite scores for the WSI, WPS, and WVI, and between these composite scores and scores on the remaining predictor measures. Intercorrelations among composite scores within the same instrument were computed to assess measurement redundancy. Finally, correlations between composite scores on different instruments were also computed to assess redundancy and to highlight the extent to which the predictors assessed similar underlying constructs. Tables of correlations appear in Appendix C.

Within-Instrument Composite Correlations

In general, correlations between composite scores from the same measure were stronger than correlations between a given measure's composite score and another measure's composite scores. This finding was due in part to overlapping content; in other words, the composite scores from the same instrument incorporated several of the same scales. The correlation between the WPS composite scores was fairly strong (r = .65), but not so high as to be considered overly redundant. The WVI composites were also strongly correlated (r = .73).

Regarding the WSI composites, it is notable that the composites targeting performance criteria were not highly correlated with the WSI composites targeting attitudinal criteria. The WSI was designed to be resistant to faking, such that if an applicant faked on some WSI dimensions, it might help inflate a performance score, but it could adversely affect scores on an attitudinal measure (e.g., Attrition Cognitions) to the extent that scores on the attitudinal measure would reflect different WSI dimensions. Thus, to the extent that the Army specifies to applicants that the Army values an applicant's standing/performance on a wide range of criteria that require high or moderate standing on various WSI dimensions, applicants may not know in which direction they should fake their responses. Thus, especially assuming tendencies to fake, it is highly desirable that several of the WSI composites had only low to moderate correlations with each other. For example, the WSI predictor for Attrition Cognitions (targeting an attitudinal criterion variable) was negatively associated with three WSI predictors targeting performance criteria: WSI Physical Fitness (r = -.42 corrected), WSI Achievement and Effort (r = -.39 corrected), and WSI Expected Future Performance (r = -.24 corrected).

In addition, when applicants indicated they would be most effective at types of work targeted toward a particular criterion variable, it did not necessarily mean that they would score highly on WSI predictors that were targeted toward other types of performance criteria. For instance, there was a relatively modest but significant correlation between the WSI predictors of General Technical Proficiency and Achievement and Effort (corrected r = .26). The WSI predictor for Teamwork yielded a significant negative correlation with the predictor for Physical Fitness and nonsignificant relationships with the WSI predictors for Expected Future Performance and Achievement and Effort.

Given the strong relations among the attitudinal variables, we expected the WSI composites to have relatively high correlations. Indeed, the WSI composites targeting attitudinal variables were correlated between an absolute value of .48 to .86; correlations with the WSI predictor for Attrition Cognitions were negative, as might be expected. The high correlations between the WSI predictors for Perceived Army Fit and Satisfaction with the Army (corrected r = .86), and between the WSI predictors for Perceived Army Fit and Career Intentions (corrected r = .75) suggest a substantial degree of overlap and redundancy between those composites. Thus, one or more of the composites may not be needed.

Between-Instrument Composite Correlations

WPS, WVI, and WSI Composite Intercorrelations

With respect to correlations between composite scores on different measures that are within the same domain, the correlations between WPS (interests) and WVI (values) were moderate (though not so high as to be considered redundant: corrected *r*s = .33 to .47).

Regarding interest and temperament composites, the WPS Subjective AFit composite was more strongly related to WSI composites than was the WPS Unit AE composite. For the work values and temperament composites, the WVI Unit ASat composite was more strongly related to the WSI composites than was the WVI Unit AE composite.

Interestingly, the WVI composite for Satisfaction with the Army was correlated moderately with the WSI composite for Satisfaction with the Army (corrected *r* = .38), though they are clearly not redundant measures. Perhaps the two composites assess different aspects of the construct domain. As such, they may increment each other's validity.

WSI Composites and Other Predictor Scales

Key relations between the 10 WSI composites and other predictor scales are listed below.

- WSI composites for Perceived Army Fit, Attrition Cognitions, Career Intentions, and Future Army Affect were independent from the other predictor measures.
- WSI Future Expected Performance was not related to the ASVAB scores, Target Tracking, or PSJT Judgment, but was related to several RBI scores.
- WSI General Technical Proficiency was related to ASVAB scores (corrected *r*s = .31 and .29 with AFQT and ASVAB Technical respectively) and Target Tracking scores (corrected *r* = .15) but was not significantly related to PSJT Judgment.
- WSI General Technical Proficiency was modestly related to several RBI scale scores, the strongest relationships being with RBI Peer Leadership (corrected *r* = .24), RBI Self Efficacy (corrected *r* = .21), and RBI Cognitive Flexibility (corrected *r* = .23).
- WSI Achievement and Effort was not related to ASVAB scores. Its strongest relations were with RBI Achievement (corrected *r* = .19), RBI Internal Locus of Control (corrected *r* = .17), and RBI Army Identification (corrected *r* = .16).
- WSI Physical Fitness was not related to ASVAB scores, Target Tracking, or PSJT Judgment. However, it was related to several RBI scales; the strongest relations were with RBI Fitness Motivation (corrected *r* = .26), RBI Army Identification (corrected *r* = .27), and RBI Achievement (corrected *r* = .19).
- WSI Teamwork was unrelated to the ASVAB scores, Target Tracking, or PSJT Judgment, and related negatively with RBI Fitness Motivation and RBI Army Identification (corrected *r*s= -.20 and -.21, respectively; higher levels of teamwork were associated with reduced motivation to stay fit or remain in the Army). One potential explanation for these somewhat unexpected negative correlations may be the

operation of another variable (e.g., gender) that is positively related to WSI Teamwork, yet negatively related to RBI Fitness Motivation and RBI Army Identification. For example, based on findings in previous chapters, females appear to have less positive affect for the Army and be lower on fitness motivation than males (see Chapters 3 and 8, respectively), but they tend to score higher than males on Teamwork performance (see Chapter 5). The differential relationships between gender and these variables might account for the negative correlations found above.

- WSI Satisfaction with the Army was negatively but weakly associated with AFQT and unrelated to Target Tracking or PSJT Judgment. The strongest relations were with RBI Fitness Motivation (corrected r = .28), RBI Achievement (corrected r = .21), and RBI Army Identification (corrected r = .33).

Interest (WPS) Composites and Other Predictor Scales

Regarding the WPS composites, Unit AE was not related to ASVAB scores or Target Tracking, whereas Subjective AFit was negatively related to AFQT (corrected r = -.25) and ASVAB Technical (corrected r = -.18). Both WPS composites were related positively with PSJT Judgment, though the relation was stronger for Unit AE. Moderate relations were observed between the WPS composites and most RBI scales. The strongest relationships for the WPS Unit AE composite were with RBI Achievement (corrected r = .45) and RBI Cognitive Flexibility (corrected r = .39). The strongest correlations for the WPS AFit composite were with RBI Achievement (corrected r = .41), RBI Army Identification (corrected r = .36), and RBI Respect for Authority (corrected r = .34).

Values (WVI) Composites and Other Predictor Scales

The two WVI composites yielded similar patterns of correlations with other predictor variables. Comparatively, the WVI Unit AE composite was more strongly related to PSJT Judgment (corrected r = .21) than was the Satisfaction with the Army composite (corrected r = .11). The WVI Unit AE composite was negatively but modestly related to the ASVAB scores and Target Tracking. The strongest relationships for the WVI Unit AE composite were with RBI Achievement (corrected r = .35) and RBI Army Identification (corrected r = .30). The strongest relations for the Unit ASat composite were with RBI Achievement (corrected r = .32), RBI Army Identification (corrected r = .41), and RBI Respect for Authority (corrected r = .26).

Composite Correlation Summary

Overall, results of the predictor cross-instrument analyses suggest little appreciable overlap between the predictors. Although some of the measures have scales that assess similar constructs and the correlations between scores on such measures were significant and moderate in strength (supporting evidence for convergent validity), the magnitude of the correlations was not so high as to suggest substantial measurement redundancy. In further support of convergent and discriminant validity, correlations between scales from different instruments indicated that scales purported to measure similar constructs were generally more strongly correlated than were scales designed to measure different constructs.

Incremental Validity of Select21 Predictor Measures

In the previous sections, we examined relations among predictor measures at both the scale-level (to assess the construct validity of the predictor measures) and the composite-level (to identify areas of overlap among predictor measures that might have implications for their operational use). In this section, we focus on the incremental validity that each Select21 predictor measure offers over the AFQT (the primary enlisted selection measure), as well as over the ASVAB Technical Composite (see Chapter 6), and the ASVAB Spatial subtest. The purpose of the latter incremental validity analyses is to assess the increment in validity of Select21 predictors over not only the current selection battery (AFQT), but also other ASVAB-based measures which could potentially be used for selection.

The incremental validity results in this chapter are presented differently than those in the instrument-specific chapters. Specifically, the focus here is on the increment in validity that each *instrument in general* provides over the AFQT and ASVAB scores. In previous chapters, we explored the incremental validity of single scales (e.g., Chapter 9, RBI), rationally-derived composite scores (e.g., Chapter 12, Psychomotor; Chapter 7, PSJT), and empirically-derived composite scores (e.g., Chapter 10, WPS; Chapter 11, WVI). Rather than trying to build or evaluate composites geared towards operational use, here we take a step back to evaluate the potential of entire instruments, with particular attention on how the instruments compare to each other with regard to predicting a given criterion. For example, how much could we increment the AFQT if we entered in all the RBI scales as additional predictors of General Technical Proficiency compared to if we entered all WVI scales? We should note that in some cases, namely incremental validity estimates for the psychomotor Target Tracking score and PSJT, the results presented in this chapter are identical to those presented in earlier chapters. However, the results in this section are presented in a different format to facilitate relative comparisons among the different predictor measures.

Another purpose of these analyses is to identify the criteria for which the Army may most benefit from identifying additional selection and classification measures to supplement the ASVAB. For example, theory and past research would suggest that one might achieve negligible validity increments over cognitive aptitude measures such as the AFQT if one is solely trying to predict "can-do" performance criteria (e.g., Core Technical Proficiency, Skill Qualifications Test scores; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Nicewander, 2003). However, as results in previous chapters highlight, when one begins to define the criterion domain more broadly to include "will-do" types of performance (e.g., Achievement and Effort, Physical Fitness) and attitudinal criteria, the potential for supplementing the ASVAB with additional predictors becomes more visible.

To facilitate these goals, we present incremental validity results organized by criterion (see Tables 13.1 and 13.2). Under each criterion, predictors are sorted in descending order according to the magnitude of their corrected incremental validity for predicting a given criterion. When interpreting these results, we focus primarily on corrected incremental validities, given that the number of scores entering into the model for each predictor varied. Specifically, only one score was entered for the PSJT and Target Tracking measures, whereas 15 scales were entered for the RBI, 15 full scores for the WSI, 28 scale-level scores for the WVI, and 14 facet-

level scores for the WPS.[52] To enable fair comparisons to be made among predictors, we used Rozeboom's (1978) shrinkage formula to account for the fact that the validity of predictors that have more elements (e.g., scale scores) in their prediction model would be expected to shrink more upon cross-validation than those with less elements.[53]

It is important to note that unlike some of the previous chapters where shrinkage formulae were not used because some of the elements in the prediction equation had already been optimized based on the criterion (see Chapter 10), none of the predictor scores included in the analyses here were optimized based on the criterion (i.e., all scores were rationally derived). Thus, in at least one respect, the corrected incremental validity results presented in this chapter allow for fairer side-by-side comparisons of the Select21 predictor measures.

Predicting Performance Criteria

In general, we found very similar patterns of results in Tables 13.1 (incremental validities over the AFQT only) and 13.2 (incremental validities over the ASVAB). The predictors that substantially incremented the validity of the AFQT, also substantially incremented the validity of the ASVAB scores when predicting a given criterion. Similarly, the relative ordering of the predictors in terms of their incremental validity remained stable within a given criterion, regardless of whether the AFQT or ASVAB scores were entered in the first step of the prediction model.

Compared to the other performance criteria, we found notably smaller levels of incremental validity when predicting General Technical Proficiency and Future Expected Performance. As alluded to above, this finding was to be expected for General Technical Proficiency, given that it appears to be the most cognitively-loaded of the performance criteria. Indeed this pattern of results is consistent with the concurrent validation results from Project A (McHenry et al., 1990, p. 346). Furthermore, definitions for each of the four dimensions underlying Future Expected Performance (see Chapter 4) suggest that the future Army will put greater cognitive demands on Soldiers. As such, one might also expect less incremental validity beyond the AFQT and ASVAB for the Future Expected Performance composite as well. Although relatively small in magnitude, several predictors did provide statistically significant increments over the AFQT and ASVAB scores when predicting General Technical Proficiency and Future Expected Performance. Most notably, the RBI provided a 17.3% gain in corrected validity ($\Delta R = .09$) over AFQT for predicting General Technical Proficiency and a 36.1% gain in corrected validity ($\Delta R = .13$) over AFQT for predicting Future Expected Performance. The WVI provided a 19.4% gain in corrected validity ($\Delta R = .07$) over AFQT for predicting Future Expected Performance.

In contrast, the Select21 predictors showed notable levels of incremental validity over the AFQT and ASVAB for predicting Achievement and Effort, Physical Fitness, and Teamwork criteria. In the case of Physical Fitness, part of the reason for the large increment of many

[52] Although the WSI comprised 16 trait statements (and as such 16 full scores), given the completely ipsative nature of the WSI full scores, one WSI score was omitted from its models to avoid complete redundancy in the set of scores for each Soldier (i.e., the sum of all 16 WSI scores for a Soldier is a constant across Soldiers). The results of the analyses would be the same regardless of which of the 16 full scores was dropped, so one was dropped at random for purposes of estimating the incremental validity of the WSI.

[53] See Chapter 6 for a general description of how Rozeboom's (1978) formula was used in this report for estimating incremental validity.

predictors was due to the fact that cognitive ability is not related to physical prowess. Based on Table 13.1 and 13.2, neither is situational judgment nor psychomotor ability. The RBI, WVI, WPS, and WSI all have scales that tap into physical fitness-related attributes (e.g., RBI—fitness motivation, WVI—valuing opportunities for physical development, WPS—interest in physical activities), and each instrument provided statistically significant and practically meaningful increments over the AFQT and ASVAB for predicting Physical Fitness performance.

Table 13.1. Incremental Validity Estimates for Select21 Predictor Measures over the AFQT

		Raw			Corrected		
Criterion/Predictor	*n*	AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
General Technical Proficiency							
RBI [15]	634	**.30**	**.44**	**.14**	.52	.60	.09
WVI [28]	700	**.30**	**.40**	**.10**	.52	.55	.03
WPS [14]	732	**.30**	**.37**	**.07**	.52	.55	.03
PSJT [1]	698	**.30**	**.33**	**.04**	.52	.54	.02
WSI [15]	645	**.30**	**.37**	**.07**	.52	.54	.02
Target Tracking [1]	724	**.30**	**.33**	**.03**	.52	.53	.02
Achievement and Effort*							
RBI [15]	497	**.16**	**.46**	**.30**	.28	.50	.22
WVI [28]	525	**.16**	**.45**	**.29**	.28	.45	.17
WPS [14]	542	**.16**	**.36**	**.20**	.28	.40	.12
PSJT [1]	698	**.15**	**.24**	**.09**	.26	.33	.07
WSI [15]	498	**.16**	**.30**	**.14**	.28	.31	.03
Target Tracking [1]	542	**.16**	**.17**	.01	.28	.27	.00
Physical Fitness							
RBI [15]	634	.00	**.37**	**.37**	.00	.32	.32
WVI [28]	700	.00	**.37**	**.37**	.00	.27	.27
WPS [14]	732	.00	**.27**	**.27**	.00	.20	.20
WSI [15]	645	.00	**.24**	**.24**	.00	.13	.13
PSJT [1]	698	.00	.05	.05	.00	.00	.00
Target Tracking [1]	724	.00	.01	.01	.00	.00	.00
Teamwork							
WPS [14]	732	.06	**.25**	**.19**	.16	.39	.23
WVI [28]	700	.06	**.26**	**.19**	.16	.36	.20
RBI [15]	634	.06	**.23**	**.17**	.16	.35	.19
PSJT [1]	698	.06	**.13**	**.07**	.16	.24	.08
WSI [15]	645	.06	.16	.10	.16	.21	.05
Target Tracking [1]	724	.06	.08	.02	.16	.17	.01
Future Expected Performance							
RBI [15]	634	**.17**	**.34**	**.17**	.36	.48	.13
WVI [28]	700	**.17**	**.32**	**.14**	.36	.43	.07
WPS [14]	732	**.17**	**.25**	**.08**	.36	.39	.03
PSJT [1]	698	**.17**	**.21**	**.04**	.36	.38	.02
WSI [15]	645	**.17**	**.24**	.07	.36	.36	.01
Target Tracking [1]	724	**.17**	**.18**	.00	.36	.35	.00

*The Criterion Situational Judgment Test (CSJT) was omitted from the Achievement and Effort composite when the PSJT was the predictor.

Table 13.1. (Continued)

Criterion/Predictor	*n*	Raw			Corrected		
		AFQT Only	AFQT + Predictor	Δ*R*	AFQT Only	AFQT + Predictor	Δ*R*
Satisfaction with the Army							
RBI [15]	630	.01	**.65**	**.63**	.02	.68	.65
WVI [28]	680	.01	**.51**	**.50**	.02	.49	.47
WPS [14]	716	.01	**.41**	**.40**	.02	.40	.38
WSI [15]	633	.01	**.35**	**.33**	.02	.31	.28
PSJT [1]	696	.01	**.29**	**.27**	.02	.30	.27
Target Tracking [1]	707	.01	.04	.03	.02	.00	.00
Perceived Army Fit							
RBI [15]	630	.00	**.74**	**.73**	.01	.81	.81
WVI [28]	680	.00	**.51**	**.51**	.01	.52	.51
WPS [14]	716	.00	**.46**	**.46**	.01	.48	.47
WSI [15]	633	.00	**.36**	**.36**	.01	.35	.34
PSJT [1]	696	.00	**.27**	**.26**	.01	.29	.28
Target Tracking [1]	707	.00	.07	.07	.01	.03	.02
Attrition Cognitions							
RBI [15]	630	**.12**	**.54**	**.42**	.23	.64	.42
WPS [14]	716	**.12**	**.43**	**.31**	.23	.51	.28
WVI [28]	680	**.12**	**.41**	**.29**	.23	.45	.23
WSI [15]	633	**.12**	**.32**	**.20**	.23	.37	.14
PSJT [1]	696	**.12**	**.24**	**.12**	.23	.33	.10
Target Tracking [1]	707	**.12**	**.16**	**.04**	.23	.25	.02
Career Intentions							
RBI [15]	630	.07	**.48**	**.41**	.11	.46	.35
WVI [28]	680	.07	**.41**	**.35**	.11	.34	.23
WPS [14]	716	.07	**.35**	**.28**	.11	.31	.20
WSI [15]	633	.07	**.30**	**.23**	.11	.23	.13
PSJT [1]	696	.07	**.15**	**.08**	.11	.16	.05
Target Tracking [1]	707	.07	.07	.00	.11	.08	.00
Future Army Affect							
RBI [15]	614	.05	**.52**	**.48**	.07	.49	.41
WPS [14]	693	.05	**.34**	**.29**	.07	.28	.21
WVI [28]	663	.05	**.34**	**.29**	.07	.19	.12
WSI [15]	619	.05	**.29**	**.24**	.07	.19	.12
PSJT [1]	675	.05	**.15**	**.11**	.07	.14	.07
Target Tracking [1]	692	.05	**.13**	**.08**	.07	.12	.04

Note. AFQT Only = Absolute correlation between the AFQT and the criterion. AFQT + Predictor = Multiple correlations (*R*) based on a regression model including the AFQT and all scores for a given predictor. Bracketed numbers are the number of scores included for each predictor. The Δ*R* column indicates the increment in estimated validity (change in *R*) obtained from adding the predictors to the AFQT. Values in the first set of columns (Raw) are based on raw data. Values in the second set of columns (Corrected) are based on correlation matrices corrected for range restriction and criterion unreliability, and *R*s that have been adjusted for shrinkage using Rozeboom's (1978) formula. Predictors are sorted in descending order of the magnitude of their corrected increment in validity over the AFQT (Corrected Δ*R*). Bolded correlations in the AFQT Only column are statistically significant ($p < .05$). Bolded values in the AFQT + Predictor column indicate that the Multiple *R* for the model with the AFQT and predictor was statistically significant ($p < .05$). Bolded values in the Δ*R* column indicate that the increment in validity was statistically significant ($p < .05$).

Table 13.2. Incremental Validity Estimates for Select21 Predictor Measures over the ASVAB

		Raw			Corrected		
Criterion/Predictor	*n*	ASVAB Only	ASVAB + Predictor	Δ*R*	ASVAB Only	ASVAB + Predictor	Δ*R*
General Technical Proficiency							
RBI [15]	470	**.34**	**.46**	**.12**	.54	.61	.07
PSJT [1]	533	**.34**	**.37**	**.03**	.54	.57	.02
WVI [28]	522	**.34**	**.43**	.09	.54	.56	.01
WPS [14]	553	**.34**	**.39**	.05	.54	.55	.01
WSI [15]	487	**.34**	**.40**	.06	.54	.55	.01
Target Tracking [1]	545	**.34**	**.35**	**.01**	.54	.55	.01
Achievement and Effort							
RBI [15]	414	**.17**	**.46**	**.29**	.26	.49	.23
WVI [28]	414	**.17**	**.46**	**.29**	.26	.43	.16
WPS [14]	414	**.17**	**.37**	**.20**	.26	.37	.11
PSJT [1]	533	**.16**	**.24**	**.09**	.25	.32	.07
WSI [15]	414	**.17**	**.31**	**.14**	.26	.29	.03
Target Tracking [1]	414	**.17**	**.17**	.00	.26	.25	.00
Physical Fitness							
RBI [15]	470	.09	**.38**	**.30**	.00	.30	.30
WVI [28]	522	.09	**.37**	**.29**	.00	.21	.21
WPS [14]	553	.09	**.28**	**.19**	.00	.16	.16
WSI [15]	487	.09	**.26**	**.18**	.00	.06	.06
PSJT [1]	533	.09	.10	.01	.00	.00	.00
Target Tracking [1]	545	.09	.09	.00	.00	.00	.00
Teamwork							
WPS [14]	553	.07	**.25**	**.18**	.13	.37	.24
RBI [15]	470	.07	**.24**	**.17**	.13	.34	.21
WVI [28]	522	.07	.26	.19	.13	.32	.19
PSJT [1]	533	.07	**.14**	**.07**	.13	.23	.10
WSI [15]	487	.07	.17	.10	.13	.16	.03
Target Tracking [1]	545	.07	.09	.02	.13	.15	.02
Future Expected Performance							
RBI [15]	470	**.20**	**.35**	**.16**	.36	.48	.12
WVI [28]	522	**.20**	**.33**	.13	.36	.41	.05
PSJT [1]	533	**.20**	**.23**	**.03**	.36	.39	.03
WPS [14]	553	**.20**	**.26**	.07	.36	.38	.01
Target Tracking [1]	545	**.20**	**.20**	.00	.36	.36	.00
WSI [15]	487	**.20**	**.26**	.06	.36	.35	.00

*The Criterion Situational Judgment Test (CSJT) was omitted from the Achievement and Effort composite when the PSJT was the predictor.

Table 13.2. (Continued)

Criterion/Predictor	*n*	Raw			Corrected		
		ASVAB Only	ASVAB + Predictor	Δ*R*	ASVAB Only	ASVAB + Predictor	Δ*R*
Satisfaction with the Army							
RBI [15]	470	.02	**.65**	**.63**	.00	.67	.67
WVI [28]	522	.02	**.51**	**.50**	.00	.46	.46
WPS [14]	536	.02	**.42**	**.40**	.00	.39	.39
PSJT [1]	533	.02	**.29**	**.27**	.00	.28	.28
WSI [15]	487	.02	**.35**	**.33**	.00	.27	.27
Target Tracking [1]	536	.02	.05	.03	.00	.00	.00
Perceived Army Fit							
RBI [15]	470	.04	**.74**	**.70**	.00	.81	.81
WVI [28]	522	.04	**.52**	**.48**	.00	.50	.50
WPS [14]	536	.04	**.46**	**.42**	.00	.46	.46
WSI [15]	487	.04	**.37**	**.33**	.00	.32	.32
PSJT [1]	533	.04	**.27**	**.23**	.00	.28	.28
Target Tracking [1]	536	.04	.08	.04	.00	.00	.00
Attrition Cognitions							
RBI [15]	470	.13	**.54**	**.41**	.20	.63	.43
WPS [14]	536	**.13**	**.43**	**.30**	.21	.50	.29
WVI [28]	522	**.13**	**.42**	**.29**	.21	.43	.23
WSI [15]	487	.13	**.32**	**.20**	.20	.34	.14
PSJT [1]	533	**.13**	**.24**	**.12**	.21	.32	.11
Target Tracking [1]	536	**.13**	**.16**	**.03**	.21	.23	.02
Career Intentions							
RBI [15]	470	.07	**.48**	**.41**	.00	.44	.44
WVI [28]	522	.07	**.41**	**.34**	.02	.29	.28
WPS [14]	536	.07	**.35**	**.28**	.02	.28	.25
WSI [15]	487	.07	**.30**	**.23**	.00	.18	.18
PSJT [1]	533	.07	**.15**	**.08**	.02	.13	.10
Target Tracking [1]	536	.07	.07	.00	.02	.00	.00
Future Army Affect							
RBI [15]	470	.10	**.53**	**.43**	.02	.47	.45
WPS [14]	522	.10	**.34**	**.24**	.04	.25	.21
WSI [15]	487	.10	**.30**	**.20**	.03	.16	.13
PSJT [1]	522	.10	**.18**	**.08**	.04	.14	.10
WVI [28]	522	.10	**.35**	**.25**	.04	.13	.09
Target Tracking [1]	522	.10	**.14**	**.04**	.04	.08	.04

Note. ASVAB Only = Multiple correlations (*R*) based on a regression model including the AFQT, ASVAB Tech Composite (see Chapter 6), and ASVAB Assembling Objects subtest. ASVAB + Predictor = Multiple correlations (*R*) based on a regression model including the aforementioned ASVAB scores and all scores for a given predictor. Bracketed numbers are the number of scores included for each predictor. The Δ*R* column indicates the increment in estimated validity (change in *R*) obtained from adding the predictors to the ASVAB scores. Values in the first set of columns (Raw) are based on raw data. Values in the second set of columns (Corrected) are based on correlation matrices corrected for range restriction and criterion unreliability, and *R*s that have been adjusted for shrinkage using Rozeboom's (1978) formula. Predictors are sorted in descending order of the magnitude of their corrected increment in validity over the ASVAB (Corrected Δ*R*). Bolded correlations in the ASVAB Only column are statistically significant ($p < .05$). Bolded values in the ASVAB + Predictor column indicate that the Multiple *R* for the model with the ASVAB and predictor was statistically significant ($p < .05$). Bolded values in the Δ*R* column indicate that the increment in validity was statistically significant ($p < .05$).

With regard to Achievement and Effort, the AFQT and ASVAB showed moderate levels of validity (.28 for AFQT, .26 for ASVAB; corrected), but those validities were significantly incremented by all the Select21 predictors (with the exception of Target Tracking). Nevertheless, the magnitude of the increment was notable only for the RBI, WVI, WPS, and PSJT. For example, addition of the RBI incremented the validity of the AFQT for predicting Achievement and Effort by 78.6% (ΔR = .22), addition of the WVI incremented it by 60.7% (ΔR =.17), addition of the WPS incremented it by 42.9% (ΔR = .12), and addition of the PSJT incremented it by 26.9% (ΔR = .07). Although the WSI significantly incremented the validity of the AFQT for predicting Achievement and Effort, the corrected value of this increment was estimated to be only .03.

Given that the Achievement and Effort composite was the performance composite that had the strongest relation to the attitudinal criteria (Chapter 5), it is possible that the incremental validity estimate for the RBI may be inflated due to inclusion of the Army Identification scale (see Chapter 9). To assess this possibility, we re-ran incremental validity analyses for the RBI without the Army Identification scale. These analyses revealed that the RBI still incremented the validity of the AFQT for predicting Achievement and Effort by 71.4% (ΔR = .20). Thus, at least for the performance criteria, inclusion of the RBI Army Idenitification scale does not appear to overly bias the estimate for the RBI's incremental validity.

Lastly, with regard to the Teamwork performance criterion, the AFQT and ASVAB showed low levels of validity (.16 for the AFQT; .13 for the ASVAB, corrected), but like Achievement and Effort, those validities were significantly incremented by all of the Select21 predictors except Target Tracking. Similar to Achievement and Effort, the WPS, WVI, RBI, and PSJT exhibited the greatest level of incremental validity. For example, the addition of the WPS incremented the validity of the AFQT for predicting Teamwork by 143.8% (ΔR = .23), addition of the WVI incremented it by 125% (ΔR = .20), addition of the RBI incremented it by 118.8% (ΔR = .19), and addition of the PSJT incremented it by 50.0% (ΔR = .08).

In general, the findings in Tables 13.1 and 13.2 are consistent with incremental validity estimates from the concurrent validation phase of Project A (McHenry et al. 1990). Specifically, in Project A, the ABLE (also a rationally-based biodata measure), emerged as the predictor with the most incremental validity for predicting non-technical proficiency criteria (i.e., Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing), followed by interest and work value-related measures. Like Project A, few experimental predictor measures provided practically meaningful increments in validity over the ASVAB for predicting the General Technical Proficiency criterion, and psychomotor ability did not appear to offer any notable increment for any of the performance criteria.[54] Taken together, these findings reinforce the importance of recognizing that the performance criterion space is multi-dimensional (Campbell, McCloy, Oppler, & Sager, 1993), and provides further construct validity evidence for the Select21 performance composites.

[54] We did observe slightly more evidence for the incremental validity of the experimental Select21 predictors for predicting the General Technical Proficiency compared to the incremental validity of the experimental Project A predictors for predicting the General Soldiering Proficiency. In Select21, the "general proficiency" criterion included ratings measures, whereas in Project A, the "general proficiency" criterion included only hands-on performance tests and job knowledge tests. Therefore, the Select21 criterion likely introduced some elements of "will-do" performance into the measure, which subsequently may have resulted in the potential for experimental predictors to increment the validity of the ASVAB in the Select21 sample.

Predicting Attitudinal Criteria

As was the case with the findings for the performance criteria, we found very similar patterns of incremental validity results for the attitudinal criteria (see Tables 13.1 and 13.2). Specifically, those predictors that substantially incremented the validity of the AFQT when predicting a given attitudinal criterion generally did as well when additional ASVAB scores were considered. Similarly, the relative predictor incremental validities remained stable within a criterion, regardless of whether the AFQT or ASVAB scores were entered in the first step of the model.

Unlike results for the performance criteria, we found consistent evidence that all of the Select21 predictor measures (except Target Tracking) significantly and meaningfully incremented the validity of the AFQT and ASVAB scores for predicting all of the attitudinal criteria. Such findings suggest that while measures of cognitive aptitude, such as the AFQT and ASVAB in general, tend not to predict attitudinal criteria, interest-based and work-values based measures do (e.g., Dawis & Lofquist, 1984; Kristof-Brown, Zimmerman, & Johnson, 2005; Tranberg, Slane, & Ekeberg, 1993). One exception to this observation worth noting is that the AFQT and ASVAB were significantly related to Attrition Cognitions. Recall from Chapter 6 that the direction of the relationship between these cognitive aptitude measures and Attrition Cognitions was significantly negative (i.e., higher aptitude Soldiers were less likely to think about breaking their enlistment contract). Despite the significant relation between Attrition Cognitions and ASVAB scores, all of the Select21 predictor measures except Target Tracking showed notable levels of incremental validity for predicting Attrition Cognitions, particularly the RBI, WPS, and WVI.

Based on the results in Tables 13.1 and 13.2, the Select21 predictor measures appear to exhibit the highest levels of incremental validity for Satisfaction with the Army and Perceived Fit with the Army. With regard to the RBI, this finding is consistent with the fact that Affective Commitment was more strongly related to these criteria than the other attitudinal criteria examined (see Chapter 3). With regard to the WVI and WPS, this finding is consistent with the hypothesis that interest-based measures and work-values based measures are more proximal to satisfaction and fit perceptions than intention-related variables such as Attrition Cognitions and Career Intentions (Dawis & Lofquist, 1984; Van Iddekinge, Putka, & Sager, 2005).

In terms of the relative performance of the Select21 predictors, the RBI always emerged as the predictor with the most incremental validity over the AFQT and ASVAB for predicting attitudes. However, as noted in Chapter 9, inclusion of the RBI Army Identification scale in the RBI predictor composite may be artificially inflating incremental validity estimates for the RBI due to predictor-attitudinal criterion item similarity. To assess this possibility, we re-ran incremental validity analyses for the RBI without the Army Identification scale. These analyses suggested that the incremental validity of the RBI for predicting attitudinal criteria drops substantially if the Army Identification scale is excluded. Nevertheless, even without this scale, the RBI still offers notable incremental validity for predicting the attitudinal criteria. For example, the RBI *with* the Army Identification scale included incremented the validity of the AFQT for predicting Satisfaction with the Army and Career Intentions by .65 and .35, respectively. In contrast, the RBI *without* the Army Identification scale included incremented the validity of the AFQT for Satisfaction with the Army and Career Intentions by .49 and .13, respectively. Thus,

excluding the Army Identification scale from the RBI results in incremental validity estimates that are notably lower than the estimates tabled above, and far closer to (and in some cases lower than) the incremental validity estimates of the other predictors (particularly the WVI and WPS).[55]

After the RBI, the measures with the next highest level of incremental validity tended to be the WVI and WPS. Although exhibiting notably lower levels of incremental validity than the RBI with the Army Identification scale included, the WVI and WPS still exhibited sizable levels of incremental validity in an absolute sense. For example, between the WVI and WPS, the minimum corrected increment in validity over the ASVAB for predicting Satisfaction with the Army and Perceived Fit with the Army was .38. After the WVI and WPS, the WSI and PSJT typically exhibited the next highest level of incremental validity over the AFQT and ASVAB for predicting attitudes. With the exception of the Future Army Affect criterion, in which the WSI exhibited levels of incremental validity that were comparable to the WVI, this relative ordering of Select21 predictor measures stayed the same across criteria.

Summary

Overall, the results of the predictor cross-instrument analyses suggest little appreciable overlap among the predictors. Although some of the measures have scales that assess similar constructs, and the correlations between these measures were significant and moderate in strength (supporting evidence for convergent validity), the magnitude of the correlations was not so high as to suggest substantial measurement redundancy. In further support of the measures' convergent and discriminant validity, correlations among scales from different instruments that purported to measure similar constructs were generally stronger than correlations with scales that were designed to measure different constructs.

In some cases, predictor scores from two instruments that were designed to assess similar constructs were not correlated as strongly as one might expect. For example, the correlation between WSI Stress Tolerance and RBI Stress Tolerance was non-significant. The content of such scales should be examined further to determine the underlying reason for this lack of association. Also, illogical patterns of correlations emerged between the WSI scales and other measures. For example, within predictor categories, the WSI scales correlated modestly, or not significantly with the other temperament measures (RBI scale scores). Furthermore, RBI scales measuring similar constructs were not associated significantly with similar to WSI scales. These results can be partially explained by the design of the WSI scales, which yield composite scores that maximally predict criterion scores (not individual temperament constructs).

In general, the pattern of incremental validities observed here is consistent with past Army research, as well as with theory underlying the predictor and criterion content domains assessed by the ASVAB and Select21 measures. Little evidence was found for the ability of the Select21 predictor measures to increment the validity of the ASVAB when predicting cognitively-laden criteria such as General Technical Proficiency and Future Expected

[55] For the record, incremental validity estimates (over the AFQT) for the RBI *without* the Army Identification scale included were as follows for the other attitudinal criteria: Perceived Army Fit (ΔR = .57, down from .81 with Army Identification included), Attrition Cognitions (ΔR = .25, down from .42), and Future Army Affect (ΔR = .19, down from .41).

Performance. On the other hand, many of the Select21 predictors showed notable levels of incremental validity over the ASVAB when predicting Achievement and Effort, Physical Fitness, and Teamwork performance. Such findings reinforce the notion that when judging the efficacy of predictors for incrementing the validity of the ASVAB, it is important to account for the multi-dimensional nature of the criterion space. Substantial levels of incremental validity were found for the RBI, WVI, and WPS for predicting the attitudinal criteria, with somewhat lower levels of validity for the WSI and PSJT. While findings for the RBI were quite strong for the attitudinal criteria, such results appeared to partially reflect criterion-related contamination stemming from the inclusion of the RBI Army Identification scale in the RBI predictor composite. Nevertheless, even with the Army Identification scale removed, the RBI still exhibited notable levels of incremental validity for predicting the attitudinal criteria.

CHAPTER 14: MILITARY OCCUPATIONAL SPECIALTY CLUSTER RESULTS

Christopher E. Sager and Huy Le
HumRRO

Overview

The purpose of this chapter is to examine the potential for Select21 predictors to be useful in enlisted job classification. As discussed in Chapter 1, however, this concurrent validation effort was not structured to address the question of the utility of the experimental predictors for classification. There were not large sample sizes or job-specific criteria for most of the military occupational specialties (MOS) included in the sample. However, as described in Chapter 2, there were reasonable numbers of participating Soldiers who could be grouped into four MOS clusters (see Table 2.4). While such a sample cannot be used to estimate the potential operational increases in predicted performance, it can be used to examine parameters that positively influence classification efficiency. In this context, classification efficiency can be viewed as the extent to which weighting predictors differently when predicting performance across MOS improves the overall predicted level of performance.[56] These parameters include statistics showing that the experimental predictors have different relations with criteria across MOS clusters. Therefore, in this chapter results are presented within each of four Select21 MOS clusters—(a) Close Combat; (b) Surveillance, Intelligence, and Communications (SINC); (c) Maintenance/Repair; and (d) Logistics/Supply. These clusters are described in Table 2.5.

The performance criteria included four Army-wide observed performance composites (i.e., General Technical Proficiency, Achievement and Effort, Physical Fitness, Teamwork) and the future oriented performance composite (i.e., Future Expected Performance).[57] Army-wide performance criteria were used because MOS-specific performance criteria were available for too few Soldiers (see Table 2.4). In addition to the five performance criteria used in the validity analyses reported in previous chapters, we included two MOS-specific scale scores from the Army Life Survey (ALS)—Perceived MOS Fit and Satisfaction with the Work Itself. We used these MOS-specific attitude scores because they were theoretically appropriate for examining the potential for classification efficiency.

Validity Estimates

A key component for a predictor's potential to contribute to classification efficiency is the extent to which its correlation with a criterion is different across jobs (Sager, Peterson, Oppler, Rosse, & Walker, 1997). Select21 generated a number of criterion scores and a large number of predictor scores. Two accommodations were made to prevent the presentation of an overwhelming number of criterion-related validity estimates in this chapter. First, for each predictor, only scores at their most specific level were used. For example, the Work Preference Survey (WPS) facet scores were used, but the WPS scale scores and optimized composite scores were not used. The

[56] While the traditional literature discusses classification efficiency in terms of maximizing performance (e.g., Sager, Peterson, Oppler, Rosse, & Walker, 1997), the concept applies equally well to maximizing positive attitudes towards work.

[57] Two Achievement and Effort composites are examined. One that includes the Criterion Situational Judgment Test (CSJT) score and one that does not. They are referred to here as Achievement and Effort (w/CSJT) and (wo/CSJT).

most specific construct relevant scores were used to maximize the opportunity to discover conceptually meaningful differences across MOS clusters. It is important to note that these scores were not always the ones that particular instrument specific chapters identified as the scores that were best for maximizing criterion-related validity estimates in the overall sample.

The second accommodation was to show results only for those predictor/criterion pairs that showed significant variation in validity estimates across MOS clusters. Table 14.1 shows the Select21 raw and corrected zero-order validity estimates for the predictors whose raw correlations differed significantly across the four MOS clusters ($p < .05$). This significance test examined the probability that each of the four MOS cluster-specific validity estimates were values sampled from the same population (Hedges & Olkin, 1985). Of the 77 predictor scores considered, 35 showed at least one such difference for at least one of the eight criteria. In this table and the remaining tables, sets of criteria and predictors are presented in the same order as they appear in the preceding predictor chapters. The bolded values in Table 14.1 do not refer to the "differences in correlations" test; they simply indicate the individual correlations that are significantly different from zero.

Examination of the validity estimates for the performance criteria show several notable results (see Table 14.1). First, as mentioned above, nearly half of the predictors (35) showed differences in validity estimates across clusters. Nevertheless, of those 35 predictors, 26 showed validity differences across clusters for only one of the eight criteria. Only three of the predictors showed validity differences for at least half of the criteria. Namely, RBI Fitness Motivation showed validity differences for six of the eight criteria, and both WPS Creativity and WSI Attention to Detail showed validity differences for four of the eight criteria. Second, the number of predictors that showed validity differences across clusters varied widely across criteria. For example, only two predictors showed validity differences across clusters for predicting General Technical Proficiency, whereas 14 predictors showed validity differences across clusters for predicting Perceived MOS Fit. Another notable result involves Achievement and Effort (both with and without the Criterion Situational Judgment Test [CSJT]), Physical Fitness, and Teamwork. For these criteria, there were a number of predictors that showed higher validity estimates for the Maintenance/Repair cluster compared to the other clusters, but these differences have no apparent explanation.

Table 14.1 shows that some predictors had substantially different validity estimates across MOS clusters for the attitude criteria. A number of the differences are straightforward to interpret, whereas others are not. For example, given the nature of Close Combat MOS, it appears reasonable that measures of Rational Biodata Inventory (RBI) Stress Tolerance, RBI Internal Locus of Control, RBI Army Identification, RBI Respect for Authority, and WPS Physical were more highly correlated with attitudinal criteria in this cluster. It also makes sense that scores on the Armed Services Vocational Aptitude Battery (ASVAB) Technical composite were the most related to Perceived MOS Fit for Maintenance/Repair Soldiers for whom knowledge and skill in the areas that this composite assesses are especially relevant to the job. However, other results are less interpretable. For example, the significant negative relationship between WSI Cultural Tolerance and Perceived MOS Fit for only the Maintenance/Repair cluster is difficult to understand. Because the WSI is an ipsative measure the negative correlation is explainable, but why its absolute value was relatively larger is less straightforward. Additionally, it is not clear why WPS Creativity had such a comparatively large negative correlation with Perceived MOS Fit for the SINC cluster.

Table 14.1. Validity Estimates by MOS Cluster

	Observed Validity Estimate				Corrected Validity Estimate			
Criterion/Predictor	CC	SINC	MR	LS	CC	SINC	MR	LS
General Technical Proficiency								
WSI: Attention to Detail	.03	**-.26**	.06	**.30**	.01	-.33	.09	.34
Psychomotor: Target Tracking	**.12**	-.01	**.29**	**.47**	.21	.04	.39	.59
Achievement and Effort (w/ CSJT)								
WSI: Attention to Detail	.06	-.19	.13	**.47**	.07	-.23	.16	.52
RBI: Cognitive Flexibility	-.01	.21	**.34**	.10	.01	.30	.44	.12
RBI: Fitness Motivation	**.17**	.02	**.35**	-.19	.19	-.03	.33	-.21
WPS: Creativity	-.07	-.08	**.27**	.23	-.08	-.07	.31	.25
WPS: High Profile	-.10	**-.25**	.00	.24	-.11	-.30	.00	.27
WPS: Mechanical	.11	-.12	**.27**	-.08	.12	-.17	.28	-.10
WPS: Physical	**.20**	.02	**.32**	-.15	.22	-.03	.34	-.17
WVI: Variety	-.07	.02	**.35**	.07	-.07	-.02	.41	.09
Achievement and Effort (w/o CSJT)								
WSI: Attention to Detail	**.11**	**-.23**	.17	**.45**	.12	-.27	.20	.50
RBI: Fitness Motivation	**.18**	-.01	**.31**	**-.24**	.22	-.04	.33	-.27
WPS: Creativity	-.04	-.09	**.23**	.13	-.04	-.09	.26	.15
WPS: Mechanical	.07	-.12	**.27**	-.07	.06	-.15	.30	-.08
WPS: Physical	**.13**	.00	**.27**	-.15	.12	-.01	.30	-.17
Psychomotor: Target Tracking	.03	**-.22**	**.22**	.13	.07	-.24	.27	.16
Physical Fitness								
ASVAB: AFQT	.06	-.17	-.02	**-.25**	.10	-.28	-.03	-.39
PSJT: Judgment	.08	-.13	**.25**	-.16	.10	-.20	.25	-.24
WSI: Self-Control	-.05	.14	**-.20**	**-.28**	-.05	.13	-.21	-.27
WSI: Cultural Tolerance	-.04	-.16	**.22**	**-.31**	-.04	-.19	.23	-.35
RBI: Interpersonal Skills-Diplomacy	.00	**.25**	**.27**	.19	.01	.29	.28	.11
RBI: Self-Esteem	.08	.08	**.29**	**.40**	.09	.09	.29	.41
WPS: Creativity	-.03	.09	.02	**.34**	-.03	.06	.02	.35
WVI: Advancement	-.04	**.25**	**-.21**	.07	-.04	.29	-.22	.03
WVI: Feedback	-.04	.10	**-.31**	-.09	-.04	.13	-.32	-.09
WVI: Influence	.05	**.21**	-.13	-.14	.06	.22	-.14	-.23
WVI: Recognition	-.02	.12	**-.39**	.01	-.02	.10	-.40	.04
WVI: Social Status	-.02	.14	**-.26**	-.17	-.02	.14	-.27	-.13
Teamwork								
WSI: Initiative	.04	-.12	**.29**	.05	.07	-.21	.49	.10
RBI: Fitness Motivation	.05	-.20	.18	-.16	.09	-.35	.29	-.24
WVI: Ability Utilization	.02	-.14	.07	**.30**	.04	-.21	.14	.51
WVI: Personal Development	-.01	-.18	.07	**.26**	-.01	-.30	.12	.45
Psychomotor: Target Tracking	-.08	**-.20**	.11	.15	-.12	-.32	.19	.30
Future Expected Performance								
WSI: Attention to Detail	.05	**-.20**	.04	**.34**	.05	-.29	.07	.47
RBI: Army Identification	.10	.10	.19	**.46**	.16	.12	.26	.62
RBI: Fitness Motivation	**.23**	-.03	**.27**	-.07	.33	-.12	.33	-.09

Table 14.1 (Continued)

Criterion/Predictor	Observed Validity Estimate				Corrected Validity Estimate			
	CC	SINC	MR	LS	CC	SINC	MR	LS
Perceived MOS Fit								
ASVAB: Technical Composite	.00	-.09	**.26**	-.12	-.04	-.19	.32	-.19
WSI: Adaptability/Flexibility	**-.15**	.09	-.18	.13	-.16	.10	-.20	.16
WSI: Cultural Tolerance	-.11	-.13	**-.31**	.20	-.11	-.16	-.33	.18
RBI: Army Identification	**.53**	**.34**	**.29**	**.32**	.55	.36	.31	.36
RBI: Fitness Motivation	**.21**	**.33**	.07	-.12	.21	.39	.05	-.14
RBI: Internal Locus of Control	**.25**	-.04	.11	-.13	.26	-.05	.15	-.16
RBI: Stress Tolerance	**.22**	.10	.09	-.17	.23	.08	.10	-.22
WPS: Creativity	-.06	**-.30**	.02	.08	-.07	-.34	.03	.09
WPS: Information Management	**-.12**	**.21**	-.13	.17	-.12	.22	-.12	.20
WPS: Lead Others	**.20**	**.31**	-.08	.14	.22	.33	-.10	.15
WPS: Physical	**.43**	**.26**	.10	.13	.46	.32	.10	.15
WPS: Work with Others	**.24**	**.20**	-.17	**.44**	.27	.26	-.19	.49
WVI: Leadership Opportunities	**.14**	**.30**	-.08	.07	.15	.34	-.08	.05
WVI: Travel	**.12**	.10	-.19	.10	.13	.12	-.21	.09
Satisfaction with Work Itself								
ASVAB: Technical Composite	**-.17**	-.19	.15	-.02	-.27	-.30	.12	-.08
RBI: Army Identification	**.48**	.20	**.29**	**.31**	.46	.21	.30	.34
RBI: Fitness Motivation	**.20**	**.34**	-.01	-.07	.18	.40	.00	-.08
RBI: Respect for Authority	**.39**	.12	**.20**	.16	.41	.09	.20	.17
WVI: Creativity	**-.16**	-.09	.18	-.10	-.22	-.12	.18	-.13

Note. $n_{\text{Close Combat}}$ = 189-352, n_{SINC} = 72-108, $n_{\text{Maintenance/Repair}}$ = 92-113, $n_{\text{Logistics/Supply}}$ = 60-82. CC = Close Combat. SINC = Surveillance, Intelligence, and Communications. MR = Maintenance/Repair. LS = Logistics/Supply. Bolded observed validity estimate are statistically significant ($p < .05$). Corrected validity estimates were corrected for measurement error in the criterion measures and range restriction due to direct selection on AFQT.

While it is true that some of the differences in validity estimates across MOS clusters are more interpretable than others, two observations are relevant. First, some scales from each predictor showed evidence of variation in criterion-related validity estimates across clusters. Second, more experimental predictors showed differences in validities across MOS clusters for criteria that reflect the will-do or motivational determinants of performance (e.g., Achievement and Effort) than criteria that depend more on can-do determinants of performance (e.g., General Technical Proficiency).

Subgroup Differences

Table 14.2 provides estimates of subgroup differences in mean scores (effect sizes) comparing MOS clusters on each of the relevant criteria. Four effect sizes were close to or greater than half of an *SD*: (a) the mean Future Expected Performance score was greater for the SINC than the Maintenance/Repair cluster, (b) the same was true for the mean Achievement and Effort (w/o CSJT) score, (c) the mean General Technical Proficiency score was greater for the SINC than the Logistics/Supply cluster, and (d) the Teamwork score was greater for the SINC than the Close Combat cluster.[58]

[58] Unfortunately, the results reported here cannot be compared to the Select21 field test results (Van Iddekinge, Sager, & Le, 2005) for which composite performance scores were not produced.

Table 14.3 provides subgroup difference estimates (effect sizes) for comparing MOS clusters on each of the relevant predictors. Generally, the Predictor Situational Judgment Test (PSJT), WSI, RBI, WPS, and Work Values Inventory (WVI) did not show substantial differences in mean scores across MOS clusters. Exceptions included a WPS Mechanical mean score that was more than three-fourths of an *SD* greater for the Maintenance/Repair cluster than the SINC and Logistics/Supply clusters. Additionally, the WPS Physical mean score for the Close Combat cluster was more than one-half of an *SD* greater than those for the SINC and Logistics/Supply clusters. Finally, the RBI Army Identification mean score for the Close Combat cluster was also more than one-half an *SD* greater than those for the SINC and Logistics/Supply clusters. On the other hand, the ability measures (i.e., AFQT, ASVAB Technical Composite, and Target Tracking) showed a number of substantial mean differences across MOS clusters. The Logistics/Supply cluster ASVAB Technical Composite mean score was at least one-half of an *SD* lower than the mean scores for the other three MOS clusters. This result was the same for Target Tracking.[59]

Mean differences on criterion and predictor scores across jobs contribute to the potential for classification efficiency (Zeidner & Johnson, 1994), and these results revealed substantial differences across the MOS clusters. SINC cluster criterion scores were greater than the Maintenance/Repair and Logistics/Supply scores for multiple measures. Predictor differences showed that the clusters differed in terms of Mechanical and Physical interests (i.e., according to WPS scores) and that the Logistics/Supply cluster differed from the other clusters on the examined ASVAB and psychomotor scores. However, this latter effect should be interpreted with caution because ASVAB subtests that contribute to the Technical composite also contribute to the operational composites that influence MOS assignment and these predictors are all positively correlated with each other. The point is that these observed differences could be partially due to range restriction on operational ASVAB composites. For example, the Maintenance/Repair cluster may have a much higher mean on the ASVAB Technical Composite than the Logistics/Supply cluster because operational ASVAB composites require higher scores on the relevant ASVAB subtests for assignment to Maintenance/Repair MOS than to Logistics/Supply MOS.

[59] These results cannot be readily compared to the Select21 field test results because subgroup differences related to MOS were not reported for predictors. Unlike the concurrent validations participants, predictor field test participants were new Soldiers who had just begun basic training and thus had not yet been involved in MOS-specific activities.

Table 14.2. Differences in Criterion Scores by MOS Cluster

							Close Combat		SINC		Maintenance/ Repair		Logistics/ Supply	
Criterion	d_{SC}	d_{MC}	d_{LC}	d_{MS}	d_{LS}	d_{LM}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
General Technical Proficiency	0.12	-0.09	**-0.40**	**-0.23**	**-0.55**	-0.33	0.01	0.54	0.07	0.45	-0.04	0.48	-0.21	0.58
Achievement and Effort (w/ CSJT)	**0.36**	-0.04	-0.08	**-0.39**	**-0.43**	-0.05	-0.01	0.48	0.16	0.53	-0.03	0.47	-0.05	0.48
Achievement and Effort (w/o CSJT)	**0.42**	-0.05	-0.02	**-0.48**	**-0.46**	0.03	-0.05	0.54	0.17	0.52	-0.08	0.52	-0.06	0.50
Physical Fitness	-0.27	-0.05	-0.32	0.20	-0.05	-0.24	0.06	0.73	-0.14	0.77	0.02	0.80	-0.18	0.84
Teamwork	**0.47**	0.12	0.20	-0.36	-0.25	0.08	-0.04	0.59	0.24	0.58	0.03	0.57	0.08	0.70
Future Expected Performance	**0.42**	-0.09	0.02	**-0.56**	**-0.44**	0.12	-0.03	0.66	0.23	0.52	-0.09	0.64	-0.02	0.63
Perceived MOS Fit	-0.23	0.07	**-0.34**	0.32	-0.11	**-0.44**	3.10	0.96	2.88	0.92	3.16	0.88	2.77	0.91
Satisfaction with Work Itself	**-0.33**	-0.01	0.02	0.33	0.34	0.03	3.09	0.89	2.80	0.87	3.08	0.88	3.10	0.93

Note. $n_{\text{Close Combat}}$ = 223-367, n_{SINC} = 84-108, $n_{\text{Maintenance/Repair}}$ = 102-115, $n_{\text{Logistics/Supply}}$ = 71-84. SINC = Surveillance, Intelligence, and Communications. d_{SC} = Effect size for SINC-Close Combat mean difference. d_{MC} = Effect size for Maintenance/Repair-Close Combat mean difference. d_{LC} = Effect size for Logistics/Supply-Close Combat mean difference. d_{MS} = Maintenance/Repair-SINC mean difference. d_{LS} = Logistics/Supply-SINC mean difference. d_{LM} = Logistics/Supply-Maintenance/Repair. Effect sizes calculated as (mean of first cluster – mean second cluster)/pooled *SD* for both clusters. Bolded effect sizes are statistically significant, $p < .05$ (two tailed).

Table 14.3. Differences in Predictor Scores by MOS Cluster

							Close Combat		SINC		Maintenance/ Repair		Logistics/ Supply	
Predictor	d_{SC}	d_{MC}	d_{LC}	d_{MS}	d_{LS}	d_{LM}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ASVAB: AFQT	0.22	0.10	**-0.40**	-0.14	**-0.71**	**-0.56**	56.56	19.02	60.64	15.84	58.40	16.60	49.22	16.33
ASVAB: Technical Composite	-0.17	0.28	**-0.87**	**0.47**	**-0.71**	**-1.06**	153.11	18.91	149.91	16.33	158.44	19.64	135.67	23.85
PSJT: Judgment	**0.29**	0.27	0.18	-0.03	-0.13	-0.10	4.54	0.39	4.66	0.38	4.65	0.34	4.61	0.33
WSI: Adaptability/Flexibility	-0.11	-0.03	-0.11	0.08	0.00	-0.08	9.10	4.34	8.61	4.19	8.96	4.32	8.59	4.86
WSI: Attention to Detail	0.05	0.25	-0.04	0.21	-0.10	-0.30	9.73	4.49	9.96	3.90	10.81	4.02	9.54	4.39
WSI: Initiative	-0.22	-0.20	-0.22	0.02	0.01	-0.01	7.81	3.60	6.98	4.22	7.06	4.01	7.03	3.72
WSI: Self-Control	**-0.41**	-0.25	-0.27	0.16	0.14	-0.02	8.42	4.30	6.69	4.16	7.36	4.08	7.26	4.10
WSI: Cultural Tolerance	0.21	-0.07	0.20	-0.29	-0.01	0.28	7.73	4.79	8.75	4.73	7.38	4.77	8.70	4.80
RBI: Army Identification	**-0.54**	-0.21	**-0.65**	0.35	-0.11	**-0.47**	3.24	0.83	2.79	0.83	3.07	0.74	2.70	0.80
RBI: Cognitive Flexibility	0.22	-0.02	-0.23	-0.23	**-0.49**	-0.21	3.40	0.71	3.55	0.69	3.39	0.76	3.24	0.59
RBI: Fitness Motivation	-0.08	0.03	-0.23	0.11	-0.13	-0.29	3.49	0.68	3.43	0.84	3.51	0.60	3.33	0.64
RBI: Internal Locus of Control	0.14	0.07	-0.14	-0.07	-0.31	-0.23	3.34	0.60	3.42	0.57	3.38	0.59	3.25	0.50
RBI: Interpersonal Skills-Diplomacy	-0.11	0.15	0.05	0.26	0.16	-0.10	3.37	0.82	3.29	0.78	3.49	0.79	3.42	0.79
RBI: Respect for Authority	0.01	0.04	-0.09	0.03	-0.10	-0.14	3.32	0.70	3.33	0.65	3.35	0.66	3.26	0.59
RBI: Self-Esteem	0.18	0.05	-0.11	-0.14	-0.32	-0.18	3.86	0.64	3.97	0.50	3.89	0.56	3.79	0.61
RBI: Stress Tolerance	-0.04	0.21	-0.09	0.26	-0.05	-0.30	2.85	0.51	2.83	0.48	2.96	0.51	2.80	0.53
WPS: Creativity	0.06	-0.01	0.09	-0.07	0.03	0.09	3.61	0.85	3.66	0.80	3.60	0.86	3.68	0.88
WPS: High Profile	0.22	-0.16	0.14	**-0.41**	-0.07	0.30	2.50	0.89	2.70	0.84	2.36	0.80	2.63	1.02
WPS: Information Management	**0.32**	0.07	**0.40**	-0.26	0.09	0.34	2.58	0.87	2.86	0.82	2.65	0.80	2.94	0.88
WPS: Lead Others	-0.21	-0.24	0.02	-0.03	0.22	0.25	3.64	0.84	3.46	0.82	3.43	0.84	3.65	0.94
WPS: Mechanical	**-0.33**	**0.42**	**-0.47**	**0.78**	-0.16	**-0.87**	3.18	1.02	2.85	0.92	3.62	1.04	2.69	1.11
WPS: Physical	**-0.91**	**-0.38**	**-0.59**	**0.52**	0.28	-0.21	3.67	0.85	2.88	0.91	3.35	0.86	3.15	0.99
WPS: Work with Others	-0.21	-0.04	0.03	0.17	0.23	0.07	3.57	0.85	3.39	0.92	3.53	0.84	3.60	0.90
WVI: Ability Utilization	**0.35**	0.22	-0.10	-0.14	**-0.48**	-0.33	0.23	1.15	0.62	1.05	0.48	1.10	0.12	1.10
WVI: Advancement	0.26	0.20	0.09	-0.06	-0.18	-0.11	0.78	1.18	1.08	0.97	1.01	1.08	0.89	1.17
WVI: Creativity	0.21	0.26	0.03	0.06	-0.19	-0.23	-0.21	1.20	0.04	0.97	0.10	1.09	-0.17	1.25
WVI: Feedback	0.29	0.11	0.12	-0.19	-0.17	0.01	-0.32	1.14	-0.02	0.86	-0.20	1.07	-0.18	1.15
WVI: Influence	0.01	-0.03	-0.02	-0.04	-0.03	0.00	-0.77	1.15	-0.77	0.88	-0.80	0.93	-0.80	1.05
WVI: Leadership Opportunities	0.03	-0.01	0.01	-0.04	-0.01	0.02	0.15	1.30	0.18	1.24	0.14	1.09	0.17	1.24
WVI: Personal Development	**0.33**	0.19	-0.05	-0.14	-0.43	-0.24	0.00	1.19	0.38	1.01	0.22	1.28	-0.06	1.07
WVI: Recognition	0.18	-0.03	0.20	-0.22	0.02	0.24	-0.09	1.28	0.13	1.13	-0.12	1.19	0.16	1.14
WVI: Social Status	0.01	-0.04	0.01	-0.05	-0.01	0.05	0.45	1.32	0.46	1.11	0.40	1.28	0.45	1.17
WVI: Travel	-0.15	-0.18	-0.27	-0.03	-0.13	-0.11	-0.98	1.39	-1.19	1.18	-1.22	1.20	-1.35	1.21

Table 14.3. (Continued)

							Close Combat		SINC		Maintenance/ Repair		Logistics/ Supply	
Predictor	d_{SC}	d_{MC}	d_{LC}	d_{MS}	d_{LS}	d_{LM}	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
WVI: Variety	-0.03	0.06	-0.01	0.10	0.02	-0.08	-0.13	1.20	-0.16	1.01	-0.05	1.21	-0.14	1.02
Psychomotor: Target Tracking	0.06	0.05	**-0.51**	-0.01	**-0.52**	**-0.54**	0.05	0.93	0.10	0.99	0.10	0.93	-0.44	1.10

Note. $n_{\text{Close Combat}}$ = 309-358, n_{SINC} = 82-114, $n_{\text{Maintenance/Repair}}$ = 104-118, $n_{\text{Logistics/Supply}}$ = 76-89. SINC = Surveillance, Intelligence, and Communications. d_{SC} = Effect size for SINC-Close Combat mean difference. d_{MC} = Effect size for Maintenance/Repair-Close Combat mean difference. d_{LC} = Effect size for Logistics/Supply-Close Combat mean difference. d_{MS} = Maintenance/Repair-SINC mean difference. d_{LS} = Logistics/Supply-SINC mean difference. d_{LM} = Logistics/Supply-Maintenance/Repair. Effect sizes calculated as (mean of first cluster – mean second cluster)/pooled *SD* for both clusters. Bolded effect sizes are statistically significant, $p < .05$ (two tailed).

Differential Prediction

Table 14.4 shows the differential prediction analysis results for each relevant predictor score. The predictors are organized by criterion in the same manner as Table 14.1. The analyses discussed here are the same as those explained in Chapter 6 for the assessment of gender, race, and ethnic group predictive bias. Table 14.4 shows three columns for each pair of MOS clusters. The first column shows the intercept differences. A negative value means that the second MOS Cluster in the pair has a higher intercept value. For example, for RBI Stress Tolerance predicting Perceived MOS Fit, the bolded intercept difference (B_{SC} = -0.27) means that Close Combat has a significantly higher intercept value than SINC. The interpretation is that a common regression formula for these clusters would be likely to underpredict fit for Close Combat Soldiers and overpredict Fit for SINC Soldiers. The second two columns show the slope for each cluster. The size of the slope represents the degree of relationship between the predictor and the criterion for Soldiers in that cluster.

Table 14.4 shows a substantial number of significant intercept differences. For example, for the SINC vs. Close Combat comparison, 41 of the possible 55 intercept differences examined were significant and there were similar results for the Maintenance/Repair vs. SINC and Logistics/Supply vs. SINC comparisons. The size and direction of these effects were consistent with the related mean differences on criterion composite scores (see Table 14.2). While the results regarding slope differences were somewhat more modest, the table does show a fair number of them. In particular, for the Maintenance/Repair vs. SINC comparison, 20 out of the possible 55 Maintenance/Repair slopes significantly favored Maintenance/Repair. This finding means that there was a stronger relationship between the relevant predictors and criteria for Maintenance/Repair Soldiers than SINC Soldiers. This effect was associated with a number of significant mean differences on the criteria and difference in validity estimates for these two clusters. Counting the number of significant values should be done with some caution because Table 14.4 shows only those criterion/predictor pairs that demonstrated variation in validity estimates across MOS clusters. These criterion/predictor pairs represent only 55 (8.9%) of the total 616 possible pairs.

Summary

As indicated earlier in this chapter, differences in validity estimates across MOS clusters and means on criteria and predictors are evidence of the potential for classification efficiency (e.g., Sager et al., 1997; Zeidner & Johnson, 1994). All eight criteria and some scales from all of the experimental predictors showed MOS cluster differences on validity estimates, means, and differential prediction analyses. While this pattern of results is not easily summarized in a concise way, a few observations are particularly noteworthy.

Table 14.4. Differential Prediction Results by MOS Cluster

	SINC vs. Close Combat			Maintenance/Repair vs. Close Combat			Logistics/Supply vs. Close Combat			Maintenance/Repair vs. SINC			Logistic/Supply vs. SINC			Logistics/Supply vs. Maintenance/Repair		
Criterion/Predictor	B_{SC}	B_S	B_C	B_{MC}	B_M	B_C	B_{LC}	B_L	B_C	B_{MS}	B_M	B_S	B_{LS}	B_L	B_S	B_{LM}	B_L	B_M
General Technical Proficiency																		
WSI: Attention to Detail	0.07	**-0.12**	0.01	-0.07	0.03	0.01	**-0.24**	**0.18**	0.01	**-0.14**	0.03	**-0.12**	**-0.31**	**0.18**	-0.12	**-0.17**	0.18	0.03
Psychomotor: Target Tracking	0.07	0.00	0.07	-0.06	0.15	0.07	-0.14	**0.24**	0.07	**-0.13**	**0.15**	0.00	**-0.21**	**0.24**	0.00	-0.07	0.24	0.15
Achievement and Effort (w/ CSJT)																		
WSI: Attention to Detail	**0.17**	-0.10	0.03	-0.03	0.06	0.03	-0.04	**0.23**	0.03	**-0.21**	0.06	**-0.10**	**-0.22**	**0.23**	-0.10	-0.01	**0.23**	0.06
RBI: Cognitive Flexibility	**0.16**	0.10	0.00	-0.01	**0.15**	0.00	-0.05	0.06	0.00	**-0.17**	0.15	0.10	**-0.21**	0.06	0.10	-0.04	0.06	0.15
RBI: Fitness Motivation	**0.19**	0.01	0.08	-0.02	0.18	0.08	-0.08	**-0.10**	0.08	**-0.20**	**0.18**	0.01	**-0.27**	-0.10	0.01	-0.07	-0.10	**0.18**
WPS: Creativity	**0.18**	-0.05	-0.03	0.01	**0.12**	-0.03	-0.03	**0.10**	-0.03	**-0.17**	**0.12**	-0.05	**-0.21**	0.10	-0.05	-0.04	0.10	0.12
WPS: High Profile	**0.19**	-0.13	-0.05	0.00	0.00	-0.05	-0.04	**0.09**	-0.05	**-0.19**	0.00	-0.13	**-0.23**	0.09	**-0.13**	-0.04	0.09	0.00
WPS: Mechanical	**0.16**	-0.07	0.05	-0.05	0.12	0.05	-0.04	-0.04	0.05	**-0.22**	**0.12**	-0.07	**-0.21**	-0.04	-0.07	0.01	-0.04	**0.12**
WPS: Physical	**0.21**	0.01	0.10	0.04	0.16	0.10	-0.03	-0.06	**0.10**	**-0.17**	**0.16**	0.01	**-0.24**	-0.06	0.01	-0.07	-0.06	**0.16**
WVI: Variety	**0.17**	0.02	-0.03	-0.03	**0.15**	-0.03	-0.05	0.04	-0.03	**-0.20**	0.15	0.02	**-0.21**	0.04	0.02	-0.02	0.04	0.15
Achievement and Effort (w/o CSJT)																		
WSI: Attention to Detail	**0.20**	**-0.13**	0.06	-0.05	0.09	0.06	-0.02	**0.22**	0.06	**-0.25**	0.09	**-0.13**	**-0.22**	**0.22**	-0.13	0.04	0.22	0.09
RBI: Fitness Motivation	**0.21**	-0.01	0.10	-0.01	0.17	0.10	-0.06	**-0.13**	0.10	**-0.21**	**0.17**	-0.01	**-0.26**	-0.13	-0.01	-0.05	-0.13	**0.17**
WPS: Creativity	**0.21**	-0.05	-0.02	0.00	**0.11**	-0.02	-0.01	0.06	-0.02	**-0.22**	**0.11**	-0.05	**-0.23**	0.06	-0.05	-0.01	0.06	0.11
WPS: Mechanical	**0.20**	-0.07	0.04	-0.07	0.13	0.04	-0.02	-0.03	0.04	**-0.27**	**0.13**	-0.07	**-0.22**	-0.03	-0.07	0.05	-0.03	**0.13**
WPS: Physical	**0.24**	0.00	0.07	0.02	0.14	0.07	-0.01	**-0.07**	0.07	**-0.21**	**0.14**	0.00	**-0.25**	-0.07	0.00	-0.03	-0.07	**0.14**
Psychomotor: Target Tracking	**0.24**	**-0.12**	0.02	-0.01	0.11	0.02	-0.02	0.06	0.02	**-0.25**	0.11	**-0.12**	**-0.26**	0.06	**-0.12**	-0.01	0.06	0.11
Physical Fitness																		
ASVAB: AFQT	**-0.17**	**-0.15**	0.04	-0.06	-0.02	0.04	**-0.36**	**-0.23**	0.04	0.12	-0.02	-0.15	-0.19	-0.23	-0.15	**-0.30**	-0.23	-0.02
PSJT: Judgment	**-0.19**	-0.10	0.06	-0.10	0.22	0.06	**-0.28**	-0.15	0.06	0.09	**0.22**	-0.10	-0.09	-0.15	-0.10	-0.18	-0.15	**0.22**
WSI: Self-Control	-0.17	0.11	-0.04	-0.07	-0.16	-0.04	**-0.26**	-0.24	-0.04	0.10	**-0.16**	0.11	-0.10	**-0.24**	0.11	-0.20	-0.24	-0.16
WSI: Cultural Tolerance	-0.17	-0.12	-0.03	-0.03	**0.17**	-0.03	**-0.22**	**-0.24**	-0.03	0.14	**0.17**	-0.12	-0.06	-0.24	-0.12	-0.20	**-0.24**	0.17
RBI: Interpersonal Skills-Diplom.	-0.17	**0.20**	0.00	-0.10	**0.20**	0.00	**-0.28**	0.16	0.00	0.07	0.20	0.20	-0.11	0.16	0.20	-0.18	0.16	0.20
RBI: Self-Esteem	**-0.22**	0.07	0.05	-0.08	**0.22**	0.05	**-0.22**	**0.33**	0.05	0.13	0.22	0.07	0.00	0.33	0.07	-0.13	0.33	0.22
WPS: Creativity	**-0.19**	0.07	-0.02	-0.06	0.02	-0.02	**-0.25**	**0.28**	-0.02	0.13	0.02	0.07	-0.06	0.28	0.07	-0.19	**0.28**	0.02
WVI: Advancement	**-0.22**	**0.23**	-0.03	-0.01	-0.16	-0.03	**-0.29**	0.06	-0.03	**0.22**	-0.16	**0.23**	-0.07	0.06	0.23	**-0.28**	0.06	-0.16
WVI: Feedback	**-0.21**	0.10	-0.03	-0.02	**-0.23**	-0.03	**-0.29**	-0.08	-0.03	0.19	**-0.23**	0.10	-0.08	-0.08	0.10	**-0.27**	-0.08	-0.23
WVI: Influence	**-0.19**	0.20	0.04	-0.02	-0.11	0.04	**-0.29**	-0.12	0.04	0.16	-0.11	**0.20**	-0.10	-0.12	**0.20**	**-0.27**	-0.12	-0.11

Table 14.4. (Continued)

	SINC vs. Close Combat			Maintenance/Repair vs. Close Combat			Logistics/Supply vs. Close Combat			Maintenance/Repair vs. SINC			Logistic/Supply vs. SINC			Logistics/Supply vs. Maintenance/Repair		
Criterion/Predictor	B_{SC}	B_S	B_C	B_{MC}	B_M	B_C	B_{LC}	B_L	B_C	B_{MS}	B_M	B_S	B_{LS}	B_L	B_S	B_{LM}	B_L	B_M
Physical Fitness (continued)																		
WVI: Recognition	**-0.20**	0.10	-0.01	-0.05	**-0.30**	-0.01	**-0.29**	0.01	-0.01	0.15	**-0.30**	0.10	-0.09	0.01	0.10	**-0.24**	0.01	**-0.30**
WVI: Social Status	**-0.19**	0.12	-0.01	-0.03	**-0.19**	-0.01	**-0.29**	-0.16	-0.01	0.16	**-0.19**	0.12	-0.10	**-0.16**	0.12	**-0.26**	-0.16	-0.19
Teamwork																		
WSI: Initiative	**0.28**	-0.06	0.02	0.12	**0.15**	0.02	0.09	0.04	0.02	-0.15	**0.15**	-0.06	-0.19	0.04	-0.06	-0.03	0.04	0.15
RBI: Fitness Motivation	**0.31**	-0.09	0.03	0.09	0.11	0.03	0.08	-0.12	0.03	**-0.22**	**0.11**	-0.09	**-0.23**	-0.12	-0.09	-0.01	**-0.12**	0.11
WVI: Ability Utilization	**0.28**	-0.09	0.01	0.09	0.04	0.01	**0.18**	**0.22**	0.01	**-0.20**	0.04	-0.09	-0.11	**0.22**	-0.09	0.09	0.22	0.04
WVI: Personal Development	**0.29**	-0.12	-0.01	0.09	0.04	-0.01	**0.16**	**0.21**	-0.01	**-0.20**	0.04	-0.12	-0.13	**0.21**	-0.12	0.07	0.21	0.04
Psychomotor: Target Tracking	**0.32**	-0.11	-0.05	0.07	0.06	-0.05	0.13	**0.10**	-0.05	**-0.25**	0.06	**-0.11**	-0.19	0.10	**-0.11**	0.06	0.10	0.06
Future Expected Performance																		
WSI: Attention to Detail	**0.27**	-0.11	0.03	-0.10	0.03	0.03	-0.03	**0.23**	0.03	**-0.36**	0.03	-0.11	**-0.30**	**0.23**	-0.11	0.07	0.23	0.03
RBI: Army Identification	**0.28**	0.05	0.06	-0.02	0.13	0.06	0.15	**0.30**	0.06	**-0.30**	0.13	0.05	-0.13	**0.30**	0.05	0.17	0.30	0.13
RBI: Fitness Motivation	**0.25**	-0.01	**0.15**	-0.04	0.19	0.15	-0.01	-0.04	**0.15**	**-0.29**	**0.19**	-0.01	**-0.27**	-0.04	-0.01	0.03	-0.04	**0.19**
Perceived MOS Fit																		
ASVAB: Technical Composite	-0.20	-0.10	0.00	0.01	**0.24**	0.00	**-0.36**	-0.09	0.00	0.21	**0.24**	-0.10	-0.16	-0.09	-0.10	**-0.37**	-0.09	**0.24**
WSI: Adaptability/Flexibility	-0.21	0.08	**-0.15**	0.07	-0.17	-0.15	**-0.28**	0.11	**-0.15**	**0.28**	-0.17	0.08	-0.07	0.11	0.08	**-0.35**	0.11	**-0.17**
WSI: Cultural Tolerance	-0.18	-0.12	-0.10	0.05	-0.29	-0.10	**-0.28**	**0.18**	-0.10	0.23	-0.29	-0.12	-0.10	**0.18**	-0.12	**-0.33**	0.18	**-0.29**
RBI: Army Identification	-0.04	0.30	0.51	0.18	0.30	**0.51**	-0.06	0.30	0.51	0.23	0.30	0.30	-0.02	0.30	0.30	-0.25	0.30	0.30
RBI: Fitness Motivation	**-0.24**	0.24	0.20	0.06	0.07	0.20	**-0.30**	-0.12	**0.20**	**0.30**	0.07	0.24	-0.06	-0.12	**0.24**	**-0.36**	-0.12	0.07
RBI: Internal Locus of Control	**-0.27**	-0.04	**0.23**	0.04	0.10	0.23	**-0.32**	-0.13	**0.23**	**0.32**	0.10	-0.04	-0.05	-0.13	-0.04	**-0.36**	-0.13	0.10
RBI: Stress Tolerance	**-0.27**	0.09	0.22	0.03	0.08	0.22	**-0.31**	-0.14	**0.22**	**0.30**	0.08	0.09	-0.04	-0.14	0.09	**-0.35**	-0.14	0.08
WPS: Creativity	-0.19	**-0.29**	-0.05	0.07	0.02	-0.05	**-0.32**	0.07	-0.05	**0.26**	0.02	**-0.29**	-0.13	0.07	**-0.29**	**-0.39**	0.07	0.02
WPS: Information Management	**-0.23**	**0.21**	-0.11	0.08	-0.13	-0.11	**-0.34**	**0.15**	-0.11	**0.32**	-0.13	**0.21**	-0.11	0.15	0.21	**-0.43**	**0.15**	-0.13
WPS: Lead Others	-0.16	0.30	0.20	0.08	-0.07	**0.20**	**-0.32**	0.12	0.20	**0.24**	-0.07	**0.30**	-0.15	0.12	0.30	**-0.40**	0.12	-0.07
WPS: Physical	0.05	0.24	**0.45**	**0.21**	0.10	**0.45**	-0.15	0.11	**0.45**	0.16	0.10	0.24	-0.21	0.11	0.24	**-0.36**	0.11	0.10
WPS: Work with Others	-0.17	0.18	0.24	0.07	-0.16	**0.24**	**-0.35**	0.38	0.24	**0.25**	-0.16	**0.18**	-0.17	0.38	0.18	**-0.42**	**0.38**	-0.16
WVI: Leadership Opportunities	**-0.24**	0.28	0.13	0.07	-0.08	0.13	**-0.33**	0.07	0.13	**0.31**	-0.08	**0.28**	-0.08	0.07	0.28	**-0.40**	0.07	-0.08
WVI: Travel	**-0.22**	0.10	0.11	0.08	**-0.19**	0.11	**-0.29**	0.10	0.11	**0.29**	**-0.19**	0.10	-0.08	0.10	0.10	**-0.37**	0.10	**-0.19**

Table 14.4. (Continued)

Criterion/Predictor	SINC vs. Close Combat			Maintenance/Repair vs. Close Combat			Logistics/Supply vs. Close Combat			Maintenance/Repair vs. SINC			Logistic/Supply vs. SINC			Logistics/Supply vs. Maintenance/Repair		
	B_{SC}	B_S	B_C	B_{MC}	B_M	B_C	B_{LC}	B_L	B_C	B_{MS}	B_M	B_S	B_{LS}	B_L	B_S	B_{LM}	B_L	B_M
Satisfaction with Work Itself																		
ASVAB: Technical Composite	**-0.29**	-0.21	-0.16	-0.04	0.13	**-0.16**	0.04	-0.02	-0.16	**0.25**	0.13	**-0.21**	**0.33**	-0.02	-0.21	0.08	-0.02	0.13
RBI: Army Identification	-0.10	0.17	**0.43**	0.11	0.28	0.43	**0.25**	0.28	0.43	0.21	0.28	0.17	**0.36**	0.28	0.17	0.15	0.28	0.28
RBI: Fitness Motivation	**-0.24**	0.24	0.18	0.00	-0.01	0.18	0.05	-0.06	**0.18**	0.24	-0.01	**0.24**	**0.29**	-0.06	**0.24**	0.04	-0.06	-0.01
RBI: Respect for Authority	**-0.27**	0.11	**0.33**	-0.02	0.19	0.33	0.06	0.15	0.33	0.25	0.19	0.11	**0.32**	0.15	0.11	0.08	0.15	0.19
WVI: Creativity	**-0.27**	-0.09	-0.15	0.02	**0.17**	-0.15	0.02	-0.08	-0.15	**0.30**	0.17	-0.09	**0.29**	-0.08	-0.09	-0.01	-0.08	0.17

Note. Regression analyses were carried out separately for each pair of MOS clusters. In each regression analysis comparing MOS Clusters, the first Cluster was coded as 1 and the second as 0 (e.g., SINC (1) vs. Close Combat (0)). SINC = Surveillance, Intelligence, and Communications. B_{SC} = Intercept difference between Surveillance, Intelligence, and Communications and Close Combat. B_{MC} = Intercept difference between Maintenance/Repair and Close Combat. B_{LC} = Intercept difference between Logistics/Supply and Close Combat. B_{MS} = Intercept difference between Maintenance/Repair and Surveillance, Intelligence, and Communications. B_{LS} = Intercept difference between Logistics/Supply and Surveillance, Intelligence, and Communications. B_{LM} = Intercept difference between Logistics/Supply and Maintenance/Repair. B_C = Slope for Close Combat. B_S = Slope for Surveillance, Intelligence, and Communications. B_M = Slope for Maintenance/Repair. B_L = Slope for Logistics/Supply. Bolded intercept differences indicate that the two MOS clusters had significant different intercepts, $p < .05$. If two slopes were significantly different from each other, the one with the largest absolute value is bolded, $p < .05$.

Six predictor measure scales showed differences in validity estimates across clusters for three or more criterion composites: (a) RBI Fitness Motivation, (b) WSI Attention to Detail, (c) WPS Creativity, (d) WPS Physical, (e) RBI Army Identification, and (f) Target Tracking (see Table 14.1). Of these predictors, WPS Physical, RBI Army Identification, and Target Tracking showed mean differences across clusters (see Table 14.3) and all showed differential prediction intercept and slope differences across clusters (see Table 14.4). Other predictors showed more targeted results focused on specific cluster comparisons or criteria. For example, when predicting performance, a number of predictors showed higher validity estimates for the Maintenance/Repair cluster compared to the others. The corrected validity estimate for WPS Work with Others was .49 for the Logistics/Supply cluster and -.19 for the Maintenance/Repair cluster when predicting the Perceived MOS Fit criterion composite. Additionally, the corrected validity estimate for Target Tracking was .59 for the Logistics/ Supply cluster and .04 for the SINC cluster when predicting General Technical Proficiency. Another salient result is the extent of mean differences on the criteria across MOS clusters. There were a number of significant mean difference estimates, and they were strongly associated with significant intercept differences in the differential prediction analyses. While mean differences on a criterion themselves do not directly affect classification efficiency, they can influence the effect that differences in validities have on classification efficiency. Depending on the size and direction of differences in validities across jobs, the number of jobs, the number of predictors being used, and other factors, mean differences on the criterion can reduce or increase potential classification efficiency. Determining the effect of criterion mean differences on classification efficiency requires a different research design than the one employed here (Zeidner & Johnson, 1994).

When interpreting these results, several considerations should be kept in mind. Only those predictor/criterion relationships that showed variation in validity estimates were shown in this chapter's tables, and only a small number (55) of the possible (616) predictor/criterion pairs are depicted. This number reflects the fact that of the 35 predictors that exhibited validity differences across clusters, 26 exhibited such differences for only one of the eight criteria. The level of job differentiation may provide some insight into this result. The analyses placed jobs into clusters and sought to differentiate between the clusters. The modest differentiation may simply underscore the difficulty of deriving meaningful clusters and changes in the levels of job description over the course of this research. Additionally, the performance measures themselves were Army-wide (i.e., not targeted to the clusters). It is possible that the use of MOS-specific performance criteria would have resulted in more evidence supporting the potential of the experimental predictors to contribute to classification efficiency. Finally, with the potential exception of RBI Fitness Motivation, which showed validity differences across clusters for six of the eight criteria, no particular predictor measure was found to be substantially superior to others in terms of the potential for improving classification efficiency. However, the evidence across the predictors and criteria suggest that there may be some potential for improvements to classification efficiency in the Army's enlisted MOS assignment process.

CHAPTER 15: SUMMARY

Teresa Russell and Deirdre Knapp (HumRRO)
Trueman Tremble (ARI)

Overview

The purpose of this final chapter is to (a) provide a brief summary of the Select21 research; (b) point out some of the innovative elements of the research, as well as the extent to which we were able to adapt to circumstances throughout the 4-year program; (c) summarize key findings (both empirical and experiential); (d) comment on the varying degrees of confidence we now have in conclusions regarding the experimental predictors and their use (e.g., we know a lot more about their potential for selection than for classification); and (e) offer suggestions for future research (some of which, as discussed later, is already underway as part of a follow-on research effort).

Research Summary

The 4-year Select21 project concerned future entry-level Soldier selection, with the goal of ensuring that the Army selects and classifies Soldiers with the knowledge, skills, and attributes (KSAs) needed to perform successfully in a transformed Army. The ultimate objectives of the project were to (a) develop and validate measures of critical attributes needed for successful execution of Future Force missions, and (b) propose use of the measures as a foundation for an entry-level selection and classification system adapted to the demands of the 21st century.

The major elements of the approach to this project were (a) future-oriented job analysis, (b) development of predictor measures suitable for predicting performance in the future Army, (c) development of performance and attitude criterion measures consistent with anticipated future Army requirements, and (d) a concurrent criterion-related validation effort. The future-oriented job analysis (Sager, Russell, Campbell, & Ford, 2005) provided the foundation for the development of new tests that could be used for recruit selection or Military Occupational Specialty (MOS) assignment/ classification (i.e., predictors) and the development of job performance measures that serve as criteria for evaluating the predictors. Development of the Select21 predictor and criterion measures was documented in Knapp, Sager, and Tremble (2005).

This report has described results of the concurrent criterion-related validation portion of the research. Additional information relevant to the validity of the Select21 measures for attitudinal criteria is presented in two reports on how well pilot and field test versions of the Select21 measures predict attrition (Putka & Bradley, 2006; Putka & Le, 2005).

Innovations and Adaptations

Future-Oriented Job Analysis

The future-oriented nature of this project required adjustments to traditional job analysis methods. One adjustment was the designation of a target future time period accompanied by

basic assumptions about the Army during this time (e.g., the simultaneous existence of forces at different stages of transformation). Additionally, we adopted a combined "top-down" and "bottom-up" approach to considering information about future projections and current performance requirements, respectively. This approach helped us combine future projections that were dynamic and relatively broad (top down), with available information about current performance requirements that was more specific (bottom up). A thorough explication of future-oriented performance requirements depended on this integration.[60] This way of looking at the future led us to include Army-wide and cluster/MOS-specific anticipated conditions in the 21st century for first-term Soldiers as a separate performance requirement product. These anticipated conditions allowed us to more fully represent broad and dynamic future projections than we would have been able to if we had restricted the analysis to more traditional performance dimensions and tasks. In fact, they were the primary input into the development of expected future performance criteria, while the Army-wide performance dimensions, Army-wide common tasks, and cluster/MOS-specific tasks were the primary input into the development of current performance criteria.

Other methodological adaptations were needed to ensure that the job analysis information would serve predictor and criterion development needs in light of both selection and classification goals. As in the Army's Project A (Campbell & Knapp, 2001), Army-wide job analysis products were designed to support the development of predictors to improve selection, while the MOS-specific products were designed to support the development of predictors to demonstrate potential improvements to classification. Descriptions of performance requirements were compiled to guide the development of criterion measures, while a list of pre-enlistment KSAs was developed to facilitate predictor development. Finally, the Select21 job analysis procedures identified future-oriented job clusters and MOS to focus on for the cluster/MOS-specific portion of the job analysis. The method identified clusters and MOS that were intended to be (a) critical to the Future Force, (b) differentiated in terms of performance requirements and pre-enlistment KSAs, and (c) practical in terms of access to sufficient subject matter experts (SMEs) to complete the job analysis and develop and evaluate predictor measures. The results of the MOS clustering and prioritization guided decisions about MOS to be included in the research program.

Measurement of Criterion Domains

Obviously, it is not possible to develop "true" future criterion measures when the future cannot be known with certainty. In Select21 and in the NCO21 research program on which it built (Knapp, Heffner, & McCloy, 2004), however, project researchers developed creative ways to integrate the best available projections about future job requirements into criterion measures that could be used with today's Soldiers. The Future Expected Performance Rating Scales and the Future Army Life Survey used the future Army conditions identified in the Select21 job analysis work to "fast-forward" respondents into the future as experts expect it to unfold. Although not possible to accomplish at the level of individual technical tasks, projecting people (Soldiers and those rating their performance) into a conceptual understanding of the future was not only feasible, but seemed to work quite well.

[60] This process was greatly facilitated by regular review of our job analysis products by the Subject Matter Expert Panel (SMEP). They were a unique set of mostly senior NCOs who, as a group, combined specific knowledge about current performance requirements with awareness of the Army's transformation efforts.

Another important methodological advance in Select21 was the inclusion of measures of attitudinal criteria along with measures of job performance. In the decades before the Services' Job Performance Measurement Projects (JPM), research often neglected the criterion, with the criterion of choice in military validation studies usually being the most available one. Consequently, the Armed Services Vocational Aptitude Battery (ASVAB) was repeatedly validated against training school grades (with considerable success) (Welsch, Kucinkas, & Curran, 1990). Of course, it is critical that the ASVAB predict training performance, and that finding in itself is noteworthy. The JPM projects, including the Army's Project A, went well beyond training validation studies, however, and showed that the ASVAB is also a very good predictor of a wide variety of job performance criteria. In Project A, job performance was conceptualized, at the broadest level, in terms of can-do and will-do facets (Campbell & Knapp, 2001), which are essentially equivalent to task and contextual performance, respectively (Borman & Motowidlo, 1993). In short, can-do aspects of job performance have been well-predicted by the ASVAB, but will-do aspects have been less so. With concerns about attrition, recent research has focused on attitudinal criteria such as attrition cognitions and satisfaction with the Army (Strickland, 2005). Select21 drew on prior research about job performance and attitudes to build a set of criterion measures that would tap both domains using a variety of measurement methods including attitude surveys, peer and supervisor ratings, a job knowledge test, a criterion situational judgment test, and personnel records. This was a significant step towards obtaining more complete coverage of criteria that are important to the Army.

Measurement of Predictor Domains

Our intent was to develop predictors that supplement the ASVAB for the prediction of performance and attitudinal criteria. Because the ASVAB predicts can-do aspects of performance well, the biggest gains in selection and classification efficiency are likely to come from the addition of measures that are not highly correlated with cognitive ability, such as measures of temperament and psychomotor abilities.

Research has repeatedly shown that measures of temperament, interests, and values are good predictors of important criteria (e.g., effort, teamwork, attrition) that are not well-predicted by the ASVAB (Campbell & Knapp, 2001). In operational or high-stakes settings, however, individuals, intentionally or not, tend to distort their responses on self-report measures so as to present themselves in a positive light. An important component of the Select21 research effort was the development of innovative measures of temperament, interests, and values that employ various methods that reduce such measures' susceptibility to response distortion.

For example, the Rational Biodata Inventory (RBI) used biodata items that, being relatively observable in contrast to items on traditional personality measures, were expected to be less fakable. Items were also selected based on their correlations with the RBI so-called "lie" scale, which was included to gauge the degree to which individual respondents appear to be misrepresenting themselves. The Work Suitability Inventory (WSI) used an innovative sorting/ranking exercise to assess temperament constructs, and sought to thwart the effects of any particular response bias by use of empirically derived scoring algorithms that differ for each criterion measure. The Predictor Situational Judgment Test (PSJT) was not developed to assess temperament per se, and we were not concerned about response distortion in this case (although we did investigate the effects of coaching on improving test scores). We did, however, attempt to

develop PSJT subscores that reflected personality traits with the idea this might be another way to capture temperament without the interference of examinee response distortion. In the end, however, our PSJT personality-based scoring key was unsuccessful.

Select21 also included self-report experimental predictors based on the concept of person-environment fit. The Work Values Inventory (WVI), for example, used a ranking exercise to determine what characteristics of work situations are particularly important to an individual (e.g., the opportunity to work with people, having clearly defined work requirements). The Work Preferences Survey (WPS) assessed an individual's work-related interests.

Prior research also suggested that psychomotor tests could supplement the ASVAB. Psychomotor tests have been shown to be good predictors of gunnery performance and certain other job performance criteria (Silva, 1997). Furthermore, classification research has suggested that psychomotor test scores are likely to enhance the classification efficiency of the ASVAB (Sager, Peterson, Oppler, Rosse, & Walker, 1997; Schmidt, Hunter, & Dunn, 1995). With these benefits in mind, we adapted psychomotor tests from Project A for use in Select21. To make the tests more portable, and perhaps more acceptable than they have been in the past, we used commercial off-the-shelf joysticks instead of a specially designed response apparatus.

In total six predictor measures were included in the concurrent validation effort—RBI, WVI, WPS, PSJT, WSI, and the psychomotor Target Tracking test.[61] As discussed below, several of these measures showed promise in Select21 for supplementing the ASVAB for the prediction of important performance and attitudinal criteria.

Validation Data Collection

Given the War on Terror, Army resources were stretched thin during the concurrent validation data collection, and we took steps to mitigate the impact of this issue on the successful completion of the research. We narrowed the scope of the concurrent validation to focus on two target MOS for job-specific criterion measurement from the six MOS originally planned. The criterion field test results also indicated that it would be sufficient to collect performance ratings from one supervisor rater rather than two, as we had originally planned. We optimized our ability to obtain this single rating by having a mail-back rating package to give to supervisors who were not able to meet with us on-site.

In securing support for the data collection, ARI requested participation by first-term enlisted Soldiers and at least one supervisor per participating first-term Soldier. The support request operationalized "first-term soldier" as a Soldier serving in his/her first term of service who had completed between 18 and 36 months time in service (TIS). The duration of initial

[61] Although not included in the concurrent validation, Select21 researchers also created a prototype measure to capture information about a range of KSAs that could be obtained through self-report of related training, experience, and credentials. The Record of Pre-Enlistment Training and Experience (REPETE) was used in the predictor field tests reported in Knapp, Sager, and Tremble (2005). Although it collected information pertinent to numerous KSAs, it emphasized the area of computer-related skills. There has long been interest in adding a computer skills related subtest to ASVAB, but the idea is hampered by the fact that tests of such skills rapidly become outdated. Finding a way to obtain verifiable information about computer skills using a strategy other than a test is a potentially important contribution.

technical training and enlistment terms vary across MOS, so this definition attempted to capture the concept of "first-term Soldier" in a way that would not be influenced by variations across MOS. It proved very difficult for the supporting installations to comply with these requirements, however, so we expanded our definition of first-term Soldier to increase the pool of eligible participants. This strategy helped improve our sample sizes and subsequent analyses suggested that our findings were not adversely affected by this decision. Specifically, correlations between predictors and criteria partialling out TIS were not very different from the comparable zero-order correlations between these variables.

The obtained data support informative conclusions about the potential value of the Select21 predictors as selection tools. Despite our efforts to adapt our strategy to the operational environment, however, we were not able to obtain sufficient sample sizes for the 25U MOS to warrant classification efficiency analyses using MOS-specific criterion data and comparing this MOS to the other target MOS (11B). As discussed in Chapter 14, however, we were able to explore the question of classification efficiency using the Army-wide criterion data and comparing results across clusters of like MOS.

Key Findings

The Criterion Domains

Five Performance Criterion Scores

Modeling exercises using scores on the performance criterion measures identified the following five job performance factors:

- General Technical Proficiency—based on the Army-Wide Job Knowledge Test (AWJKT) score, the Weapons Qualification score, and peer and supervisor ratings of Common Task Performance, MOS-Specific Task Performance, Communication, Information Management, Problem Solving, and Adaptation.
- Achievement and Effort—included prior military education and disciplinary actions, the Criterion Situational Judgment Test (CSJT) score, and peer and supervisor ratings of Effort and Initiative, Professionalism/Personal Discipline, and Personal/Professional Development.
- Physical Fitness—based on the Army Physical Fitness Test (APFT) score and peer and supervisor ratings of Physical Fitness.
- Teamwork—made up of peer and supervisor ratings of Supports Peers and Exhibits Tolerance rating scales.
- Future Expected Performance—based on peer and supervisor ratings of expected performance in four different anticipated future conditions: Individual Pace and Intensity, Learning Environment, Disciplined Initiative, and Communication Method and Frequency.

These performance factors appear quite similar to those found in Project A (Campbell & Knapp, 2001). For example, like Project A, the Select21 performance model included factors for

General Technical Proficiency (similar to the General Soldiering Proficiency factor in the five-factor model of first term performance in Project A), Achievement and Effort (similar to the Effort and Leadership factor in Project A), and a Physical Fitness factor. Although these factors were similar to those found in Project A, they were not identical. For example, unlike Project A, we were unable to find evidence for an MOS-specific Core Technical Proficiency factor. The lack of evidence for such a factor in the Select21 may simply reflect the fact that Project A included MOS-specific hands-on job samples and a larger sample of job knowledge tests.[62] Another difference between the Select21 results and the first term Project A results is that no evidence emerged in Select21 that differentiated a Personal Discipline factor from the Achievement and Effort factor. That is, rather than appearing as a separate factor as in Project A, Disciplinary Actions appeared in Select21 as a negative indicator of Achievement and Effort.

All of the Select21 performance composites demonstrated adequate discriminant validity, and most appear to be reasonably reliable. The estimated reliabilities of the Teamwork (.35) and Future Expected Performance (.54) composites were quite low, however, particularly given that they reflect the average across multiple raters (i.e., they are not single-rater reliability estimates). The low reliabilities of the composites can be traced back to the low interrater reliability found for individual performance dimensions that underlie these composites. Despite their limitations, these two criteria were important enough to retain.

Attitudinal Criterion Scores

There were a large number of scale scores yielded by the Army Life Survey (ALS) and Future Army Life Survey (FALS). Empirical approaches did not help reduce the attitudinal criterion "space." Accordingly, we used a rational approach to select a subset of the scales for predictor validation analyses. We chose scales to meet three objectives: (a) representation of current and future-oriented constructs, (b) balance in terms of the proximity of the chosen scales to the Select21 predictors and actual attrition and re-enlistment behavior, and (c) ready interpretability to those without a background in psychology. Toward those ends, we selected five attitudinal scales on which to focus for the validation effort:

- Satisfaction with the Army—a 10-item scale from the Army Life Survey (ALS) that focuses on Soldiers' satisfaction with Army life in general.
- Perceived Army Fit—a 6-item scale from the ALS that assesses how well Soldiers perceive themselves as fitting in the Army in general.
- Attrition Cognitions—a 3-item scale from the ALS assessing the degree to which Soldiers have thought of leaving the Army.
- Career Intentions—a 5-item scale from the ALS assessing Soldiers' intentions to re-enlist and make the Army a career.
- Future Army Affect—a 5-item scale from the Future Army Life Survey (FALS) assessing the extent to which Soldiers have positive feelings about expected future Army conditions.

[62] As discussed in Chapter 1, MOS-specific job-knowledge tests were available for some, but not most, Soldiers in the Select21 sample.

Overall, the psychometric properties of these attitudinal scales were good. All scales exhibited sufficient levels of variance and had acceptable levels of internal consistency. Correlations among the scales were moderate, suggesting that they were conceptually distinct. It is important to bear in mind, however, that this is only a subset of the scale scores used in analyses.

Performance and Attitudinal Criterion Correlations

The pattern of relations between performance and attitudinal criteria revealed some findings of note. Two attitudinal criteria, Career Intentions and Future Army Affect, were generally unrelated to any of the performance criteria. In contrast, satisfaction with the Army, Attrition Cognitions, and Perceived Army Fit were significantly related to almost all of the performance criteria (average r = .17 for Satisfaction with the Army, -.25 for Attrition Cognitions, and .25 for Perceived Army Fit), indicating that Soldiers who are satisfied with the Army, perceive that they fit well with the Army, or have few thoughts of attriting tend to score higher on all of the performance composites[63].

Of the various performance criteria, the Achievement and Effort criterion tended to correlate most highly with all of the attitudinal criteria. Conceptually, this makes sense. Soldiers with positive attitudes toward the Army are likely to be more motivated, and will likely receive higher scores on Achievement and Effort (a will-do criterion that is a function of motivation).

Current versus Future Criteria

Results suggested that we were somewhat successful in developing measures that distinguished current performance and attitudes from future-oriented performance and attitudes. Regarding performance criteria, modeling analyses supported a general future performance factor underlying the AW FX rating scales in addition to the four current performance factors. With regard to attitudes, the FALS scales exhibited only small to moderate correlations with the ALS scales, indicating that Soldiers' attitudes toward the future Army were not simply a function of their attitudes about the current Army.

Validation: Improving Selection and Classification

Consistent with prior research, scores on the ASVAB continued to be good predictors of can-do performance criteria and to have less validity for predicting will-do and attitudinal criteria. AFQT, Spatial, and Technical scores from the ASVAB yielded significant correlations with General Technical Proficiency, Achievement and Effort, and Future Expected Performance scores. It is important to note that the prediction of future expected performance is a new finding, and one that bears emphasis. The ASVAB scores were not strong predictors of Physical Fitness and Teamwork performance. ASVAB scores yielded small, but significant correlations with

[63] The opposing results regarding separation intentions (i.e., Attrition Cognitions versus Career Intentions) may at first seem counterintuitive, but are likely linked to the relationships observed with the ASVAB (see Chapter 6). That is, poorer performers (who also tend to score lower on the ASVAB) are more likely to consider breaking their enlistment contract, whereas better performers evidently understand the negative consequences associated with attrition and thus decide to honor their enlistment contract despite their desire to leave the Army—which they probably plan to do following the completion of their initial enlistment term.

Attrition Cognitions; thus, higher ASVAB scores appeared to be somewhat related to having fewer thoughts about leaving the Army prior to the end of the enlistment contract.

Improving Prediction of Performance Criteria

The ASVAB is such a good predictor of General Technical Proficiency (AFQT corrected $r = .52$, ASVAB corrected $r = .54$), that it is difficult to find predictors that increment its prediction in this arena. Even so, several predictors did provide small, but statistically significant, increments in validity over the AFQT and ASVAB scores for predicting General Technical Proficiency, most notably the RBI.

ASVAB scores also predicted Achievement and Effort (AFQT corrected $r = .28$, ASVAB corrected $r = .26$) and Future Expected Performance (corrected *r*s for both AFQT and the full ASVAB were .36) to a lesser magnitude, leaving greater room for improvement. Here, the RBI, WVI, WPS, PSJT, and WSI scores all added significantly to the validity of ASVAB scores for predicting Achievement and Effort, with the ΔR ranging from .23 for the RBI to .03 for the WSI. The RBI and PSJT scores added to ASVAB validity for predicting Future Expected Performance.

ASVAB scores did not significantly predict either the Teamwork or the Physical Fitness performance criteria, though the RBI and WPS scores added significantly to the prediction of both. In addition, the PSJT score incremented the prediction of Teamwork, and the WVI and WSI scores added to the prediction of Physical Fitness.

Improving Prediction of Attitudinal Criteria

As we both expected and hoped, all of the Select21 predictor measures (except Target Tracking) significantly and meaningfully incremented the validity of the AFQT and ASVAB scores for predicting all of the attitudinal criteria. In particular, the RBI, WVI, and WPS consistently yielded significant corrected/adjusted incremental validities of .20 or more for predicting current attitudes. The WPS and RBI also incremented validity over ASVAB for predicting future attitudes by .20 or more. These findings confirmed prior research indicating that measures of cognitive aptitude tend not to be predictive of the general attitudes examined here, whereas interest-based and work-values based measures do tend to be predictive of such attitudes (e.g., Dawis & Lofquist, 1984; Kristof-Brown, Zimmerman, & Johnson, 2005; Tranberg, Slane, & Ekeberg, 1993). One exception in Select21 is that the AFQT and ASVAB scores yielded small, but significant, negative correlations with Attrition Cognitions. Soldiers scoring higher on these cognitive aptitude measures were less likely to think about breaking their enlistment contract.

Improving Prediction with Select21 Predictor Scales and Empirical Keys

There were interesting validity results at the scale or subscore level for many of the predictors; those results were described in earlier chapters. It is also important to note that empirical keying is highly desirable for some of the Select21 predictors, but the estimated validities summarized in this section were not based on such empirical keys. For example, the WSI uses an innovative ranking procedure to minimize faking, but the procedure results in

ipsative scores. McCloy and Putka (Chapter 8 of this report) have devised an empirical keying approach that can be used to maximize prediction of a chosen criterion while minimizing problems associated with ipsativity. Additional research validating and cross-validating the results could benefit the WSI and perhaps other Select21 measures.

Improving Fairness

Supplements to the ASVAB could affect the fairness of the Army's selection and classification decisions (as defined by professional standards [SIOP, 2003]). For the prediction of General Technical Proficiency, ASVAB scores showed little or no differential prediction, and when it occurred, it showed overprediction of Black Soldiers' performance. However, when ASVAB scores were used to predict performance criteria that are likely to be a function of non-cognitive variables such as motivation and personality (e.g., Achievement and Effort, Teamwork, Future Expected Performance), significant underprediction of females' performance was more likely to occur. Combining the ASVAB scores with non-cognitive (i.e., personality and other) variables in the prediction equation could (a) increase validity and (b) decrease differential prediction for these criteria.

Improving MOS Classification

The Select21 concurrent validation sample could not provide the basis for directly evaluating the potential utility of the experimental predictors for supporting classification of enlisted personnel. Sample sizes were relatively small, and we did not collect MOS-specific job performance criteria for most of the MOS. Even so, we did obtain sufficient predictor and Army-wide criterion data for subgroup analyses at the MOS cluster level for four MOS clusters—(a) Close Combat; (b) Surveillance, Intelligence, and Communications (SINC); (c) Maintenance/Repair; and (d) Logistics/Supply.

Several Select21 predictors showed promise for increasing classification efficiency, even without the benefit of MOS-specific criteria. Six predictor scales yielded differences in validity estimates across clusters for three or more criterion composites: (a) RBI Fitness Motivation, (b) WSI Attention to Detail, (c) WPS Creativity, (d) WPS Physical, (e) RBI Army Identification, and (f) Target Tracking. Other predictors showed more targeted results. For example, the corrected validity estimate for the WPS Work with Others scale was .49 for the Logistics/Supply cluster and -.19 for the Maintenance/Repair cluster when predicting the Perceived MOS Fit criterion composite. Additionally, the corrected validity estimate for Target Tracking was .59 for the Logistics/ Supply cluster and .04 for the SINC cluster when predicting General Technical Proficiency. Out of any predictor examined, the RBI Fitness Motivation scale showed perhaps the most potential for increasing classification efficiency in that it showed validity differences across clusters for six of the eight criterion measures considered.

Generalizability of Research

As with any piece of research, there are limitations to the generalizability of inferences that can be made based on findings from a local validation effort. Here we discuss characteristics of the Select21 research sample and research design that limit the extent to

which we can assume the Select21 findings generalize to an operational Army pre-enlistment test context.

First, the Select21 sample does not exactly mirror the population of first-term Soldiers. Approximately 54% of the sample was from the Close Combat MOS cluster. The other three MOS clusters—SINC, Maintenance/Repair, and Logistics/Supply—made up less than half of the sample (i.e., the sample size for each of the other clusters was less than 150). In comparison, roughly 26% of Army active duty enlisted members were in infantry and related jobs in 2004 (*Population Representation in the Military Services*: http://www.dod.mil/prhome/poprep2004/).

A similar limitation has to do with sample sizes, regardless of the proportional representation. For example, the total sample had 83 females compared to 728 males. This proportion (i.e., 10% female and 90% male) is not too disparate from the 2004 Army enlisted population distribution, i.e., 15% female; 85% male (*Population Representation in the Military Services*: http://www.dod.mil/prhome/poprep2004/). However, the total sample of 84 females was so small that for some analyses, data were available for very few females. We are very confident in our results with regard to the Close Combat cluster and for males, where sample sizes were large; however, results for smaller MOS clusters and females are likely to be less stable.

The concurrent validation research design fundamentally differs from an operational setting in which predictors would be administered to applicants instead of experienced Soldiers. There are several ways in which one might expect findings from a concurrent design to differ from an operational setting, but here we focus on two factors that are of particular concern in the Select21 research—(a) the response distortion that is likely to occur when non-cognitive measures such as the RBI, WPS, and WVI are administered to applicants, and (b) contaminate variation in predictor measures arising from their administration to incumbents.

With regard to the response distortion issue, Soldiers participating in a research effort have little motivation to make themselves look appealing to the Army in their responses to experimental measures. Not only will respondents be more motivated to look good in an operational setting, one can expect at least some applicants to be coached on how to do well on the pre-enlistment screening tests. The extent to which the effectiveness of the Select21 self-report temperament and interest measures would be compromised in an operational environment needs to be addressed using a research design that more closely resembles an operational setting.

Another factor that may affect the generalizablity of the concurrent validation results is that Soldiers' responses to predictor measures may be influenced by the experiences they have gained in the Army. For example, many of the items on the RBI ask about past behavior, but for experienced Soldiers, this includes post-enlistment behaviors likely influenced by the fact they have been in the Army. In applicant samples, respondents can only answer RBI items based on "pre-Army" behavior. Another example of this phenomenon occurs with the WSI where Soldiers were asked what types of work they think they would be able to perform best. Their answers may be influenced by their Army experience. A question that is difficult to answer for both the RBI and WSI (and to a lesser extent the other Select21 non-cognitive measures) is that had respondents completed such measures based solely on their pre-Army experiences, would it significantly affect the validity estimates observed in the Select21 concurrent sample. As was the

case with response distortion, this is an issue that could be addressed by future efforts that examine the performance of the Select21 non-cognitive measures in an applicant setting.

Foundation for Follow-On Research

ARI has embarked on follow-on research to Select21, as *Investigations into Army Enlisted Classification Systems (Army Class)* is designed to pick up where Select21 left off. Concurrent validation data are being collected from Soldiers in five MOS (11B, 19K, 25U, 63B, 68W/91W) and will be combined with the data collected from 11B and 25U Soldiers in Select21. The Army Class criteria include MOS-specific job knowledge tests, and the plan is to use these data to get a better estimate of the classification efficiency of the experimental predictors.

The Army Class concurrent validation will be followed by a longitudinal validation. This longitudinal validation is expected to be the capstone to the entire research program, as it will be the most challenging test of how well the surviving predictors can be expected to work upon operational implementation.

Future Research Directions

Several years ago, the Army and the Air Force jointly sponsored a project to define a joint-service selection and classification research agenda. We revisited that agenda to identify areas still needing research attention today, and we added several areas that have emerged since that time.

Criterion Policy

> "An organization's choice of criteria for personnel research significantly affects how research results will influence the design of the selection and classification system. In effect, criterion policy reflects the organization's intended definition for effective performance in that organization, and the types of predictors that are used in selection and classification decision making will depend upon the criteria against which they are compared. Systematic consideration of criterion policy is necessary so that informed decisions can be made about future predictor and criterion development" (Campbell, Russell, & Knapp, 1994).

Findings in the Project A, NCO21 (Knapp et al., 2004), and Select21 projects all confirm that the criterion matters, as validation results differ substantially by the criterion of choice. By default or by design, the Army's use of ASVAB classification composites validated against Skill Qualification Test (SQT) scores reflects a policy that seems to imply that MOS Technical Proficiency is the most important (if not only) criterion that needs to be predicted by selection and classification personnel tests.[64] Future criterion policy issues facing the Army have to do with how job performance (or more broadly, organizational fit) is defined, measured, and used in the selection and classification context. Is there a consensus within the Army about the goals of criterion measurement? Should non-technical aspects of job performance such as the individual's effort and achievement or ability to work with a team play a more important role in selection and

[64] SQT scores can be thought of as measures of MOS Technical Proficiency.

classification decisions? Research on criterion policy, conducted with policy makers, could be used to develop or identify consensus. Moreover, decisions by policy makers about the merits of various criteria could be used to guide funding and resource allocations.

Selection and Classification Algorithms

Once a decision is made to include multiple criteria in selection and classification research, how will it be implemented? How can the maximum potential gain from classification be estimated given that there are choices among predictor batteries, performance goals, and criterion measurement methods? There are a large number of permutations of predictors, criteria, and goals. How can we efficiently simulate the outcomes of different predictor/criterion/goal combinations? How successfully can operational job assignment procedures capture the potential classification gains?

In Closing

This report has focused on the results of the Select21 concurrent validation. Earlier project reports described the job analysis (Sager et al., 2005) and measure development work (Knapp, Sager, & Tremble, 2005) that led to this stage. Companion reports examined the extent to which pilot and field test versions of the Select21 measures predict attrition (Putka & Le, 2005; Putka & Bradley, 2006). In a final Select21 project report (Knapp, Tremble, Russell, & Sellman, 2007), we attempt to integrate the Select21 work, prior research efforts (e.g., the NCO21 research program), and work currently underway (i.e., the Army Class research program) to see where this path is taking the Army in terms of a strong foundation for improved enlisted Soldier selection and classification that meets its future needs.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Assouline, M., & Meir, E. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior, 31*, 319-332.

Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types. *Personnel Psychology, 56*, 45-74.

Bartlett, C.J., Bobko, P., Mosier, S.B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology, 31,* 233-241.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Manual for the differential aptitude tests: Forms S and T (5th ed.).* New York: The Psychological Corporation.

Borman, W.C., & Motowidlo, S.J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W.C. Borman (Eds.) *Personnel selection in organizations.* San Francisco: Jossey-Bass Publishers.

Borman, W.C., Kubisiak, U.C., & Schneider, R.J. (1999). Work styles. In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R. Jeanneret, & E.A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 213-226). Washington DC: American Psychological Association.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Cable, D. M., & Edwards, J. R. (2004). Complementary and supplementary fit: A theoretical and empirical integration. *Journal of Applied Psychology*, *89*(5), 822-834.

Camp, C. C., & Chartrand, J. M. (1992). A comparison and evaluation of interest congruence indeces. *Journal of Vocational Behavior, 41*, 162-182.

Campbell, J. P., & Johnson, J. W. (1992). Results of the second-tour validation (CVII). In J. P. Campbell & L. M. Zook (Eds.) *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1991 Fiscal Year* (Technical Report 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J.P., Hanson, M. A., & Oppler, S. H. (2001). Modeling performance in a population of jobs. In J. P. Campbell & D. J. Knapp, D.J. (Eds.), *Exploring the limits in personnel selection and classification* (pp. 307-333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt, W.C. Borman, & Associates (Eds.) *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass Publishers.

Campbell, J.P., Russell, T.L., & Knapp, D.J. (1994). *Roadmap: An agenda for joint-service classification research* (AL/HR-TP-1994-0003). Brooks AFB, TX: Human Resources Directorate Manpower and Personnel Research Division.

Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology, 65*, 407-414.

Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115-124.

Clifton, T.C., Kilcullen, R.N., Reiter-Palmon, R., & Mumford, M.D. (1992, August). *Comparing different background data scaling procedures using triple cross validation*. Paper presented at the 100th annual meeting of the American Psychological Association, Washington DC.

Collins, M., Le, H., & Schantz, L. (2005). Job knowledge criterion tests. In D.J. Knapp, C.E. Sager, & T. R. Tremble (Eds.) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (ARI Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Costa, P.T., Jr., Terracciano, A., & McCrae, R.R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322-331.

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353-379). Stanford, CA: Stanford University Press.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*, 456-473.

Cullen, M. J., Sackett, P. R., & Lievens, F. (2004, April). *Threats to the operational use of situational judgment tests in the college admission process*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis: University of Minnesota Press.

Dunbar, S.B., & Novick, M.R. (1988). On predicting success in training for men and women: Examples from Marine Corps clerical specialties. *Journal of Applied Psychology, 73,* 545-550.

Edwards, J. R. (1991). Person-job fit: A conceptual integration, literature review and methodological critique. *International review of industrial/organizational psychology* (Vol. 6, pp. 283-357). London: Wiley.

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology, 46*, 641-665.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fleishman, E.A. (1967). Performance assessment based on an empirically derived task taxonomy. *Human Factors, 9,* 349-366.

Ford, L. A., Campbell, R. C., Campbell, J. P., Knapp, D. J, & Walker, C. B. (2000). *21st Century soldiers and noncommissioned officers: Critical predictors of performance* (Technical Report 1102). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Gamache, L.M., & Novick, M.R. (1985). Choice of variables and gender differentiated prediction within selected academic programs. *Journal of Educational Measurement, 22,* 53-70.

Gay, E. G., Weiss, D. J., Hendel, D. D., Dawis, R. V., & Lofquist, L. H. (1971). *Minnesota studies in vocational rehabilitation: xxviii: Manual for the Minnesota Importance Questionnaire.* Minneapolis: University of Minnesota.

Grafton, F.C., Czarnolewski, M.Y., & Smith, E.P. (1988). *Relationship between Project A psychomotor and spatial tests and TOW gunnery performance: A preliminary investigation* (SCTA Working Paper RS-WP-87-100). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*, 463-488.

Guilford, J. P., & Lacey, J. I. (1947). *Printed Classification Tests, A.A.F., aviation psychological progress research report, 5.* Washington, DC: U.S. Government Printing Office.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hesketh, B., & Gardner, D. (1993). Person-environment fit models: A reconceptualization and empirical test. *Journal of Vocational Behavior, 42*, 315-332.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.

Hogan, J., & Hogan, R. (1996). *Motives, values, and preferences inventory manual*. Tulsa, OK: Hogan Assessment Systems.

Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hom, P. W., & Griffeth, R. W. (1995). *Employee turnover*. Cincinnati, OH: South-Western.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Humphreys, L.G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology, 71* (2), 327-333.

HumRRO (2006). *Investigations into Army Enlisted Classification Systems Research Execution Plan* (Interim Report 06-65). Alexandria, VA: Human Resources Research Organization.

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72-98.

Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage Publications.

Hunter, J.E., Schmidt, F.L., & Rauchenberger, J.M. (1977). Fairness of psychological tests: Implications of four definitions for selection utility and minority hiring. *Journal of Applied Psychology, 62,* 245-260.

Jackson, D.N., Wrobleski, V.R., & Ashton, M.C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance, 13*, 371-388.

Jex, S. M. (1998). *Stress and job performance: Theory, research and implications for managerial practice*. Thousand Oaks, CA: Sage Publications.

Judge, T.A., Higgins, C. A., Thoresen, C. J., & Barrick, M.R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52,* 621-652.

Kass, R. A., Mitchell, K. J., Grafton, F. C., & Wing, H. (1983). Factorial validity of the Armed Services Vocational Aptitude Battery (ASVAB), Forms 8, 9, and 10: 1981 Army applicant sample. *Educational and Psychological Measurement, 43*, 1077-1087.

Kay, G., Dolgin, D., Wasel, B., Langelier, M., & Hoffman, C. (1999). *Identification of the cognitive, psychomotor, and psychosocial skill demands of uninhabited combat aerial vehicle (UCAV) operators*. Patuxent River, MD: Naval Air Systems Command.

Keenan, P.A., Russell, T.L., Le, H., Katkowski, D., & Knapp, D.J. (2005). Performance rating scales. In D.J. Knapp, C.E. Sager, & T. R. Tremble (Eds.) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (ARI Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., Goodwin, G.F., Chen, G., Wisecarver,M.M., & Sanders, M.G. (in press) Psychological predictors of special forces leader performance, *Military Psychology*.

Kilcullen, R.N., Goodwin, J, Chen, G., Wisecarver, M., & Sanders, M. (2002). *Identifying agile and versatile officers to serve in the objective force*. Paper presented at the 2002 Army Science Conference, Orlando, FL.

Kilcullen, R.N., Mael, F.A., Goodwin, G.F., & Zazanis, M.M. (1999). Predicting U.S. Army Special Forces field performance. *Human Performance in Extreme Environments, 4(1),* 53-63.

Kilcullen, R. N., Putka, D. J., McCloy, R. A., & Van Iddekinge, C. H. (2005). Development of the Rational Biodata Inventory (RBI). In D.J. Knapp, C.E. Sager, & T. R. Tremble (Eds.) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (ARI Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kilcullen, R.N., White, L.A., Mumford, M.D., & Mack, H. (1995). Assessing the construct validity of rational biodata scales. *Military Psychology, 7(1),* 17-28.

Kilcullen, R.N., White, L.A., Sanders, M., & Hazlett, G. (2003). *The Assessment of Right Conduct (ARC) administrator's manual.* (Research Note 2003-09). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., Burnfield, J.L., Sager, C.E., Waugh, G.W., Campbell, J.P. Reeve, C.L. Campbell, R.C., White, L.A., & Heffner, T.S. (2002). *Development of predictor and criterion measures for the NCO21 research program* (Technical Report 1128). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Carter, G. W. (2003). *Examination of job satisfaction in Army enlisted personnel: Its antecedents and consequences.* Unpublished manuscript.

Knapp, D.J., McCloy, R.A., & Heffner, T.S. (Eds.) (2004). *Validation of measures designed to maximize 21st-century Army NCO performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Sager, C. E., & Tremble, T. R. (Eds.). (2005) *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., Tremble, T.R., Russell, T.L., & Sellman, W.S. (2007). *Future-Oriented Army Enlisted Personnel Selection and Classification Project (Select21) summary report* (FR-06-79). Alexandria, VA: Human Resources Research Organization.

Knapp, D.J., Waters, B.K., & Heggestad, E.D. (Eds.) (2002). *Investigations related to the implementation of the Assessment of Individual Motivation (AIM)* (Study Note 2002-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurements, and implications. *Personnel Psychology, 49*, 1-49.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*, 281-342.

Kulka, R. A. (1979). Interaction as person-environment fit. In L. R. Kahle (Ed.), *New directions for methodology of behavioral science* (pp. 55-71). San Francisco: Jossey-Bass.

Lievens, F., & Sackett, P.R. (2006). *Multidimensional predictors in high-stakes testing: Issues and strategies with creating alternate forms*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Linacre, J.M. (1997). KR-20 / Cronbach alpha or Rasch reliability: Which tells the "truth"? *Rasch Measurement Transactions, 11*, 580–581. Retrieved April 13, 2006, from http://www.rasch.org/rmt/rmt113l.htm.

Linn, R.L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 139-161.

Linn, R.L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A.K. Wigdor & W.R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies, Part II: Documentation Section*. Washington, DC: National Academy Press.

Linn, R.L. (1994). Fair test use: Research and policy. In M.G. Rumsey, C.B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification*. Hillsdale, NJ: Lawrence Erlbaum.

Linz, S., & Semykina, A. (in press). Gender differences in personality and earnings: Evidence from Russia. *Journal of Economic Psychology*.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *Journal of Social Psychology*, *137*, 369-373.

McCloy, R.A., Campbell, J.P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493-505.

McCloy, R.A., & Putka, D.J. (2005). The Work Suitability Inventory. In D.J. Knapp, C.E. Sager, & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McCloy, R.A., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999). *Development of the O*NET computerized work importance profiler*. Raleigh, NC: National Center for O*NET Development.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9,* 103-113.

McHenry, J. J., Hough, L. M., Toquam, J.L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335-354.

McHenry, J.J. & Rose, S.R. (1986). *The validity and potential usefulness of psychomotor ability tests for personnel selection and classification* (Research Note 88-13). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Meyer, J. P., & Allen, N. J. (1991). The measurement and antecedent of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology, 63*, 1-18.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8*, 161-168.

Moss, F. A., & Hunt, T. (1926). Ability to get along with others. *Industrial Psychology,* 170-178.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation*. Journal of Applied Psychology, 75*, 640-647.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557-576.

Mumford, M.D., & Stokes, G.S. (1991). Developmental determinants of individual action: Theory and practice in the application of background data. In M.D. Dunnette (Ed.) *The handbook of industrial and organizational psychology* (2nd ed.). Orlando, FL.: Consulting Psychologists Press.

Mumford, M.D., Uhlman, C.E., & Kilcullen, R.N. (1992). The structure of life history: Implications for the construct validity of background data scales. *Human Performance, 5*, 109-137.

Nicewander, A. (2003). Some considerations in the search for new measures for predicting success in military occupations. In U.S. Army Research Institute for the Behavioral and Social Sciences (Ed.), *Proceedings of the ARI Workshops on New Directions in Selection and Classification*. (pp. 434-444). Alexandria, VA: Editor.

North, R. A., & Griffin, G. R. (1977). *Aviator selection 1919-1977* (NAMRL Special Report 77-2; AD-A048105). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Office of the Under Secretary of Defense, Personnel and Readiness. (2004) *Population representation in the military services*. Retrieved May 11, 2006, from http://new.humrro.org/poprep04/appendix/appendix.html.

Oppler, S.H., McCloy, R.A., & Campbell, J.P. (2001). The prediction of supervisory and leadership performance. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 389-409). Mahwah, NJ: Lawrence Erlbaum Associates.

Oppler, S.H., McCloy, R.A., Peterson, N.G., Russell, T.L., & Campbell, J.P. (2001). The prediction of multiple components of entry-level performance. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 349-388). Mahwah, NJ: Lawrence Erlbaum Associates.

Oppler, S. H., Peterson, N. G., & Rose, A. M. (1996). Basic validation results for the LVII sample. In J. P. Campbell & L. M. Zook (Eds.) *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel, Annual Report, 1993 Fiscal Year* (Research Note 96-73). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Oppler, S.H., Russell, T.L., Rosse, R.L., Keil, C., Meiman, E.P., & Welsh, J.R. (1997). *Item evaluation for ASVAB technical subtests project: Final Report.* Washington, DC: American Institutes for Research.

O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-job fit. *Academy of Management Journal, 34*, 487-516.

Owens, W.A. (1976). Background data. In M.D. Dunnette (Ed.), *Handbook of industrial psychology*. Chicago: Rand-McNally.

Palmer, P., Hartke, D.D., Ree, M. J., Welsh, J.R., & Valentine, L. (1988). *Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11)* (Technical Paper 87-48). Brooks Air Force Base: Air Force Human Resource Laboratory.

Peterson, N.G. (Ed.) (1987). *Development and field test of the Trial Battery for Project A* (Technical Report 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N.G., Russell, T.L., Hallam, G. Hough, L., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1992). Analysis of the experimental predictor battery. In. J.P. Campbell & L.M. Zook (Eds.), *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel: Annual Report 1990 FY* (Research Note 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary theory and practice* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Putka, D. J. (2005, April). Using spline regression models to examine person-environment fit. In C. E. Sager (Chair), *Assessing Person-Environment Fit for Selection*. Symposium conducted at the 20th Annual Society for Industrial and Organizational Psychology Conference, Los Angeles, CA.

Putka, D. J., & Bradley, K. (2006). *Relations between Select21 predictor measures and first-term attrition* (IR-06-44). Alexandria, VA: Human Resources Research Organization.

Putka, D.J., & Campbell, R.C. (2005). Knapp, D.J., McCloy, R.A., & Heffner, T.S. (Eds.), *Validation of measures designed to maximize 21st-century Army NCO performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Putka, D. J., & Le, H. (2005). *Select21 attrition update: Initial entry training attrition* (Interim Report 05-58). Alexandria, VA: Human Resources Research Organization.

Quintela, J. (2003). *The effects of motivational fit on employee job performance, work attitudes and intent to leave*. Unpublished doctoral dissertation. The Ohio State University.

Ramsberger, P. F., Wetzel, E. S., Sipes, D. E., & Tiggle, R. B. (1999). *An assessment of the values of new recruits* (FR-WATSD-99-16). Alexandria, VA: Human Resources Research Organization.

Ree M. J., & Carretta, T. R. (1995). *Correlation of general cognitive ability and psychomotor tracking tests* (AL/HR-TP-1995-0009). Brooks Air Force Base, TX: Human Resources Directorate, Armstrong Laboratory.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology, 44,* 321-332.

Ree, M. J., Mullins, C. J., Matthews, J. J., & Massey, R. H. (1982). *Armed Services Vocational Aptitude Battery: Item and factor analysis of Forms 8, 9, and 10* (AFHRL-TR-81-55). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Reilly, R.R., & Chao, G.T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1-62.

Riegelhaupt, B.J., Harris, C.D., & Sadacca, R. (1987). *The development of administrative measures as indicators of soldier effectiveness* (Technical Report 754). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Roberts, H.E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in Officer Training School selection decisions. *Military Psychology, 8* (2), 95-113.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin, 85*, 1348-1351.

Russell, T.L., Katkowski, D., Le, H., & Rosse, R. L. (2005). Psychomotor tests. In D. J. Knapp, C.E. Sager, & T. R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T.L., Le, H., & Knapp, D.J. (2005). Record of Pre-Enlistment Training and Experience (REPETE). In D. J. Knapp, C.E. Sager, & T. R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T.L., Peterson, N.G., Rosse, R.L., Toquam-Hatten, J., McHenry, J.J., & Houston, J.S. (2001). The measurement of cognitive, perceptual, and psychomotor abilities. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum Associates.

Russell, T.L., Reynolds, D.H., & Campbell, J.P. (1994). *Building joint-service classification research roadmap: Individual differences measurement* (AL/HR-TP-1994-0009). Brooks Air Force Base: Armstrong Laboratory.

Sackett, P. R., & Mavor, A. (Eds.) (2002). *Attitudes, aptitudes, and aspirations of American youth: Implications for military recruitment.* Washington, D.C: National Academies Press.

Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects for a post-affirmative-action world. *American Psychologist*, *56*(4), 302-318.

Sager, C. E., Peterson, N. G., Oppler, S. H., Rosse, R. L., & Walker, C. B. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology, 9*, 97-120.

Sager, C. E., Russell, T. L., Campbell, R. C., & Ford, L. A. (2005). *Future soldiers: Analysis of entry-level performance requirements and their predictors* (Technical Report 1169). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schmidt, F. L., Hunter, J. E., & Dunn, W. L. (1995). *Potential utility increases from adding new tests tot the Armed Services Vocational Aptitude Battery (ASVAB)* (NPRDC-TN-95-5). San Diego, CA: Navy Personnel Research and Development Center.

Schmidt, F.L., Pearlman, K., & Hunter, J.E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Personnel Psychology, 33*, 705-724.

Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.

Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J.A. Weekley, & R.E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application.* Mahwah, NJ: Lawrence Erlbaum Associates.

Schmitt, N., Gooding, R. Z., Noe, R. D., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 407-422.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*, 437-453.

Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior, 17, 639-655.*

Schoenfeldt, L.F. (1989). *Biographical data as the new frontier in employee selection research.* Address presented at the annual meeting of the Division of Evaluation, Measurement and Statistics of the American Psychological Association, New Orleans.

Schratz, M. K., & Ree, M. J. (1989). Enlisted selection and classification: Advances in testing. In M. F. Wiskoff & G. M. Rampton (Eds.), *Military personnel measurement: Testing, assignment, evaluation* (pp. 1 - 40). New York: Praeger.

Schwab, D.P., & Oliver, R.L. (1974). Predicting tenure with biographical data: Exhuming buried evidence. *Personnel Psychology, 27*, 125-128.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues, 50*, 19-45.

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*, 956-970.

Silva, J. M. (1997). *Using psychomotor ability for selecting TOW Gunners* (Technical Report 1059). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Sisson, D. (1948). Forced-choice: The new Army rating. *Personnel Psychology, 1*, 365-381.

Smith, E.P., & Graham, S.E. (1987). *Validation of psychomotor and perceptual predictors of Armor Officer M-1 gunnery performance* (Technical Report 766). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Smith, E.P., & Walker, M.R. (1987). Testing psychomotor and spatial abilities to improve selection of TOW gunners. *Proceedings of the Military Testing Association Meeting*.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH: Author.

Spector, P. E. (1997). *Job satisfaction: Application, assessment, cause, and consequences.* Thousand Oaks, CA: Sage Publications.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041-1053.

Street, D. R., & Dolgin, D. L. (1994). *Computer-based psychomotor tests in optimal training track assignment of student naval aviators* (NAMRL-1391). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Strickland, W. J. (Ed.) (2005). *A longitudinal study of first term attrition and reenlistment among FY1999 enlisted accessions* (Technical Report 1172). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Thorndike, R. L. (1936). Factor analysis of social and abstract intelligence. *The Journal of Educational Psychology,* XXVII, 231-233.

Thorndike, R. I. (1949). *Personnel selection: Test and measurement techniques.* New York: Wiley.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Tinsley, H. E. A. (2000). The congruence myth: An analysis of the efficacy of the person-environment fit model. *Journal of Vocational Behavior, 56*, 147-179.

Toquam, J., Peterson, N., Rosse, R., Ashworth, S., Hanson, M., & Hallam, G. (March, 1986). *Concurrent validity data analyses: Cognitive paper-and pencil and computer-administered predictors*. A paper presented at the Project A IPR & SAG Meeting in Alexandria, VA.

Tranberg, M., Slane, S., & Ekeberg, S. E. (1993). The relation between interest congruence and satisfaction: A meta-analysis. *Journal of Vocational Behavior, 42*, 253-264.

Uhlman, C.E., Reiter-Palmon, R., & Connelly, M.S. (1990). *A comparison and integration of empirical keying and rational scaling of biographical data items*. Paper presented at the Southeastern Psychological Association.

U.S. Army. (2001). *Concepts for the Objective Force, United States Army white paper*. Online at: http://www.army.mil/features/WhitePaper/default.htm.

U.S. Army. (2002, December). *Objective Force in 2015 white paper*. Arlington, VA: Department of the Army: Objective Force Task Force.Verquer, M. L., Beehr, T. A., & Wagner, S. H. (2003). A meta-analysis of relations between person-

U.S. Department of Defense (1982). *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery (ASVAB)*. Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics).

Van Iddekinge, C. H., Putka, D. J., & Sager, C. E. (2005). Person-environment fit measures. In D. J. Knapp, C. E. Sager, & T. R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Van Iddekinge, C. H., Sager, C. E., & Le, H. (2005). Cross instrument analyses. In D.J. Knapp, C.E. Sager, & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557-574.

Wainer, D., & Thissen, H. (2001). True score theory: The traditional method. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 44-47). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Walker, C.B. (1985). *The fakability of the Army's Military Applicant Profile (MAP)*. Paper presented at the Association of Human Resources Management and Organizational Behavior proceedings, Denver, CO.

Wall, J.E., Wise, L.L., & Baker, H.E. (1996). Development of the Interest-Finder – A new RIASEC-based interest inventory. *Measurement and Evaluation in Counseling and Development, 29*, 134-152.

Waugh, G.W., & Russell, T. L. (2005). Criterion situational judgment test (CSJT). In D. J. Knapp, C. E. Sager, & T. R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and performance criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700

Welsh, J. R., Jr., Kucinkas, S. K., & Curran, L. T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies* (AFHRL-TR-90-22). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Welsh, J. R., Jr., Watson, T. W., & Ree, M. J. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Predicting military criteria from general and specific abilities* (AFHRL-TR-90-63). Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

White, L.A., & Kilcullen, R.N. (1992). *The validity of rational biodata scales*. Paper presented at the 100th meeting of the American Psychological Association, Washington DC.

White, L.A., & Young, M.C. (1998). *Development and validation of the Assessment of Individual Motivation (AIM)*. Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, CA.

White, L.A., Young, M.C., & Rumsey, M.G. (2001). ABLE implementation issues and related research. In J.P. Campbell & D.J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 525-558). Mahwah, NJ: Lawrence Erlbaum Associates.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.

Wright, S.S., & Miederhoff, P.A. (1999). Selecting students with personal characteristics relevant to pharmaceutical care. *American Journal of Pharmaceutical Education, 63*, 132-138.

Zeidner, J., & Johnson, C.D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C.B. Walker, & J.B. Harris (Eds.), *Personnel selection and classification* (pp. 377-410), Hillsdale, N.J.: Erlbaum.

APPENDIX A

MEAN CORRELATIONS UNDERLYING THE FINAL PERFORMANCE MODEL

Table A1. Mean Correlations Underlying the Final Performance Model

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Army-Wide Job Knowledge Test		.00	.03	.01	.04	.01	.03	.01	.03	.01	.03	.01	.03	.01
2	PFF Weapons Qualification	.17		.03	.01	.03	.01	.03	.01	.03	.01	.03	.01	.04	.01
3	COPRS Common Task Performance (Peer)	.06	.01		.04	.03	.03	.04	.04	.03	.03	.04	.04	.03	.03
4	COPRS Common Task Performance (Supv)	.04	.10	.26		.03	.01	.04	.01	.04	.01	.04	.01	.04	.01
5	COPRS MOS-Specific Task Performance (Peer)	.00	.07	.63	.25		.03	.04	.03	.04	.03	.03	.03	.03	.03
6	COPRS MOS-Specific Task Performance (Supv)	.06	.05	.26	.66	.27		.03	.01	.03	.01	.03	.01	.04	.01
7	COPRS Communication (Peer)	.04	.04	.51	.20	.49	.18		.03	.04	.03	.03	.03	.04	.03
8	COPRS Communication (Supv)	.07	.05	.16	.55	.18	.47	.21		.04	.01	.04	.01	.03	.01
9	COPRS Adaptation (Peer)	.06	.02	.48	.19	.40	.17	.38	.14		.04	.04	.04	.03	.03
10	COPRS Adaptation (Supv)	.02	.03	.20	.62	.20	.54	.15	.40	.15		.04	.01	.04	.01
11	COPRS Information Management (Peer)	.02	.02	.49	.21	.46	.18	.53	.14	.44	.13		.03	.03	.03
12	COPRS Information Management (Supv)	.04	.03	.23	.64	.24	.58	.22	.52	.18	.57	.17		.04	.01
13	COPRS Problem Solving (Peer)	.10	.00	.54	.21	.50	.20	.49	.16	.51	.17	.54	.21		.04
14	COPRS Problem Solving (Supv)	.04	.06	.19	.60	.20	.57	.15	.53	.13	.57	.12	.63	.17	
15	PFF Military Education	.04	.04	.14	.20	.10	.13	.09	.08	.09	.13	.07	.14	.10	.15
16	COPRS Efforts and Initiative (Peer)	.02	-.05	.51	.19	.49	.21	.35	.10	.41	.12	.49	.13	.48	.14
17	COPRS Efforts and Initiative (Supv)	.10	.03	.23	.64	.19	.63	.18	.53	.17	.59	.21	.58	.21	.61
18	COPRS Professionalism &Personal Discipline (Peer)	.00	-.03	.47	.24	.44	.22	.39	.17	.42	.16	.40	.20	.44	.18
19	COPRS Professionalism &Personal Discipline (Supv)	.01	-.03	.19	.63	.20	.53	.14	.52	.15	.60	.15	.55	.20	.59
20	PFF Army Physical Fitness Test	.01	.01	.05	.07	.01	-.01	.03	-.03	.02	.07	.05	.02	.03	.01
21	COPRS Physical Fitness (Peer)	-.07	-.05	.34	.12	.34	.11	.25	.03	.28	.11	.26	.06	.32	.01
22	COPRS Physical Fitness (Supv)	-.04	.02	.13	.39	.15	.33	.12	.33	.06	.35	.09	.31	.11	.32
23	COPRS Personal & Professional Development (Peer)	.01	.00	.52	.22	.51	.17	.47	.14	.39	.15	.48	.17	.42	.14
24	COPRS Personal & Professional Development (Supv)	-.01	.03	.25	.66	.24	.60	.20	.54	.16	.58	.20	.57	.22	.57
25	Criterion Situational Judgment Test	.21	-.15	.09	.09	.06	.08	.09	.15	.03	.11	.06	.11	.07	.11
26	COPRS Support Peers (Peer)	-.06	-.08	.42	.14	.37	.14	.34	.07	.33	.11	.33	.12	.34	.11
27	COPRS Support Peers (Supv)	-.01	-.08	.21	.51	.18	.47	.17	.44	.14	.50	.18	.51	.21	.45
28	COPRS Exhibits Tolerance (Peer)	.02	-.07	.30	.08	.27	.10	.31	.03	.30	.02	.32	.04	.31	.01
29	COPRS Exhibits Tolerance (Supv)	-.04	.03	.12	.43	.11	.38	.16	.30	.06	.40	.12	.40	.11	.29
30	PFF Deviance	.00	-.01	-.12	-.21	-.11	-.15	-.12	-.24	-.08	-.17	-.05	-.19	-.07	-.18
31	FX Individual Pace and Intensity (Peer)	.05	.03	.55	.23	.50	.16	.40	.15	.43	.15	.47	.17	.48	.15
32	FX Individual Pace and Intensity (Supv)	.07	.05	.29	.65	.26	.59	.20	.51	.20	.58	.20	.58	.26	.57
33	FX Learning Environment (Peer)	.07	.00	.51	.15	.49	.12	.44	.13	.40	.09	.48	.15	.44	.12
34	FX Learning Environment (Supv)	.14	.06	.23	.61	.24	.55	.23	.55	.16	.54	.15	.61	.20	.58
35	FX Disciplined Initiative (Peer)	.07	.02	.53	.21	.47	.17	.47	.19	.45	.12	.44	.19	.47	.16
36	FX Disciplined Initiative (Supv)	.05	.05	.28	.64	.26	.59	.24	.48	.17	.53	.21	.60	.22	.59
37	FX Communication Method and Frequency (Peer)	.06	.01	.50	.23	.47	.19	.53	.15	.43	.19	.48	.22	.45	.17
38	FX Communication Method and Frequency (Supv)	.13	.09	.28	.64	.22	.55	.22	.52	.21	.55	.18	.53	.27	.50

Note. Values below the diagonal are means of the correlations across 500 random datasets (n = 370). Values above the diagonal are standard deviations of the correlations across 500 datasets.

Table A1. (Cont.)

		15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Army-Wide Job Knowledge Test	.00	.03	.01	.03	.01	.00	.03	.01	.03	.01	.00	.04	.01	.04
2	PFF Weapons Qualification	.00	.03	.01	.03	.01	.00	.03	.01	.03	.01	.00	.03	.01	.03
3	COPRS Common Task Performance (Peer)	.03	.04	.04	.03	.04	.04	.04	.04	.03	.03	.03	.04	.03	.04
4	COPRS Common Task Performance (Supv)	.01	.04	.01	.03	.01	.01	.03	.01	.04	.01	.01	.04	.01	.04
5	COPRS MOS-Specific Task Performance (Peer)	.02	.03	.03	.03	.03	.03	.04	.03	.03	.03	.03	.04	.03	.04
6	COPRS MOS-Specific Task Performance (Supv)	.01	.03	.01	.03	.01	.01	.03	.01	.03	.01	.01	.03	.01	.03
7	COPRS Communication (Peer)	.04	.04	.03	.04	.03	.03	.04	.03	.03	.03	.03	.04	.03	.04
8	COPRS Communication (Supv)	.01	.04	.01	.03	.01	.01	.04	.01	.03	.01	.01	.04	.01	.04
9	COPRS Adaptation (Peer)	.02	.04	.04	.03	.04	.04	.04	.04	.04	.04	.03	.04	.04	.04
10	COPRS Adaptation (Supv)	.01	.04	.01	.03	.01	.01	.03	.01	.03	.01	.01	.04	.01	.04
11	COPRS Information Management (Peer)	.03	.03	.03	.03	.04	.03	.04	.03	.03	.03	.03	.04	.04	.04
12	COPRS Information Management (Supv)	.01	.03	.01	.03	.01	.01	.03	.01	.03	.01	.01	.04	.01	.04
13	COPRS Problem Solving (Peer)	.03	.03	.04	.03	.04	.04	.04	.03	.03	.04	.03	.04	.04	.03
14	COPRS Problem Solving (Supv)	.01	.03	.01	.03	.01	.01	.03	.01	.03	.01	.01	.03	.01	.04
15	PFF Military Education		.03	.01	.03	.01	.00	.02	.01	.03	.00	.00	.04	.01	.02
16	COPRS Efforts and Initiative (Peer)	.08		.04	.03	.04	.04	.03	.03	.03	.04	.03	.04	.03	.04
17	COPRS Efforts and Initiative (Supv)	.13	.17		.04	.01	.01	.03	.01	.03	.01	.01	.04	.01	.03
18	COPRS Professionalism/Personal Discipline (Peer)	.10	.59	.18		.03	.03	.04	.03	.03	.03	.03	.04	.03	.04
19	COPRS Professionalism/Personal Discipline (Supv)	.13	.23	.69	.29		.01	.03	.01	.03	.01	.01	.04	.01	.04
20	PFF Army Physical Fitness Test	-.04	-.05	.05	-.02	.04		.03	.01	.03	.01	.00	.04	.01	.03
21	COPRS Physical Fitness (Peer)	.03	.38	.10	.35	.11	.31		.03	.03	.03	.03	.04	.03	.04
22	COPRS Physical Fitness (Supv)	.05	.09	.40	.10	.39	.30	.33		.03	.01	.01	.03	.01	.04
23	COPRS Personal/Professional Development (Peer)	.10	.55	.19	.54	.22	.02	.40	.17		.03	.03	.03	.03	.04
24	COPRS Personal/ Professional Development (Supv)	.16	.22	.64	.26	.62	.07	.13	.42	.27		.01	.04	.01	.04
25	Criterion Situational Judgment Test	.05	.09	.10	.10	.10	.03	.05	.02	.10	.08		.04	.01	.03
26	COPRS Support Peers (Peer)	.03	.47	.10	.48	.13	-.04	.28	.02	.41	.12	.02		.04	.04
27	COPRS Support Peers (Supv)	.13	.20	.56	.21	.63	-.09	.03	.17	.17	.52	.09	.12		.03
28	COPRS Exhibits Tolerance (Peer)	.02	.28	.06	.32	.11	-.01	.20	.00	.33	.10	.01	.39	.10	
29	COPRS Exhibits Tolerance (Supv)	.10	.12	.34	.14	.39	.01	.01	.18	.12	.39	.04	.11	.58	.09
30	PFF Deviance	.00	-.09	-.23	-.18	-.32	-.10	-.09	-.19	-.16	-.26	-.08	-.04	-.20	-.04
31	FX Individual Pace and Intensity (Peer)	.10	.49	.18	.44	.19	.09	.42	.19	.56	.24	.04	.34	.15	.22
32	FX Individual Pace and Intensity (Supv)	.14	.19	.59	.19	.57	.08	.16	.43	.23	.61	.10	.11	.48	.04
33	FX Learning Environment (Peer)	.06	.49	.12	.45	.16	.03	.31	.08	.51	.16	.05	.40	.13	.31
34	FX Learning Environment (Supv)	.11	.11	.51	.17	.53	.07	.08	.38	.20	.58	.12	.10	.45	.07
35	FX Disciplined Initiative (Peer)	.07	.47	.19	.55	.23	.02	.34	.12	.57	.27	.11	.37	.15	.34
36	FX Disciplined Initiative (Supv)	.15	.20	.60	.21	.57	.04	.12	.39	.23	.63	.09	.14	.45	.07
37	FX Communication Method and Frequency (Peer)	.10	.41	.19	.49	.17	-.01	.28	.08	.52	.22	.07	.37	.18	.28
38	FX Communication Method and Frequency (Supv)	.09	.17	.54	.21	.51	.08	.13	.37	.21	.58	.10	.13	.42	.07

Note. Values below the diagonal are means of the correlations across 500 random datasets (n = 370). Values above the diagonal are standard deviations of the correlations across 500 datasets.

Table A1. (Cont.)

		29	30	31	32	33	34	35	36	37	38
1	Army-Wide Job Knowledge Test	.01	.00	.03	.01	.03	.01	.03	.01	.03	.01
2	PFF Weapons Qualification	.01	.00	.03	.01	.03	.01	.03	.01	.03	.02
3	COPRS Common Task Performance (Peer)	.03	.03	.03	.04	.04	.03	.03	.03	.03	.04
4	COPRS Common Task Performance (Supv)	.01	.01	.04	.01	.04	.01	.04	.01	.04	.01
5	COPRS MOS-Specific Task Performance (Peer)	.03	.04	.03	.03	.04	.03	.04	.03	.03	.04
6	COPRS MOS-Specific Task Performance (Supv)	.01	.01	.03	.01	.03	.01	.03	.01	.03	.01
7	COPRS Communication (Peer)	.04	.04	.04	.03	.04	.03	.03	.03	.03	.04
8	COPRS Communication (Supv)	.01	.01	.04	.01	.04	.01	.04	.01	.04	.01
9	COPRS Adaptation (Peer)	.04	.04	.04	.04	.04	.04	.04	.03	.04	.04
10	COPRS Adaptation (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.03	.01
11	COPRS Information Management (Peer)	.04	.03	.03	.04	.03	.04	.04	.03	.03	.04
12	COPRS Information Management (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.04	.01
13	COPRS Problem Solving (Peer)	.04	.04	.03	.04	.03	.04	.04	.04	.04	.04
14	COPRS Problem Solving (Supv)	.01	.01	.03	.01	.04	.01	.04	.01	.04	.01
15	PFF Military Education	.01	.00	.02	.01	.03	.01	.03	.01	.02	.01
16	COPRS Efforts and Initiative (Peer)	.04	.04	.03	.03	.03	.04	.03	.03	.03	.04
17	COPRS Efforts and Initiative (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.03	.01
18	COPRS Professionalism &Personal Discipline (Peer)	.04	.04	.04	.03	.03	.03	.03	.03	.03	.04
19	COPRS Professionalism &Personal Discipline (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.04	.01
20	PFF Army Physical Fitness Test	.01	.00	.03	.02	.04	.01	.04	.01	.03	.02
21	COPRS Physical Fitness (Peer)	.03	.03	.03	.03	.04	.03	.03	.03	.04	.03
22	COPRS Physical Fitness (Supv)	.01	.01	.04	.01	.04	.01	.03	.01	.03	.01
23	COPRS Personal & Professional Development (Peer)	.04	.04	.03	.04	.03	.03	.03	.03	.04	.04
24	COPRS Personal & Professional Development (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.03	.01
25	Criterion Situational Judgment Test	.01	.00	.03	.02	.03	.01	.03	.01	.03	.02
26	COPRS Support Peers (Peer)	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
27	COPRS Support Peers (Supv)	.01	.01	.03	.01	.04	.01	.03	.01	.03	.01
28	COPRS Exhibits Tolerance (Peer)	.03	.04	.04	.04	.04	.04	.04	.04	.04	.04
29	COPRS Exhibits Tolerance (Supv)		.01	.04	.01	.04	.01	.04	.01	.04	.01
30	PFF Deviance	-.11		.04	.01	.04	.01	.04	.01	.04	.01
31	FX Individual Pace and Intensity (Peer)	.10	-.14		.03	.03	.03	.03	.03	.04	.04
32	FX Individual Pace and Intensity (Supv)	.33	-.27	.24		.04	.01	.03	.00	.04	.01
33	FX Learning Environment (Peer)	.07	-.07	.64	.18		.04	.03	.04	.03	.04
34	FX Learning Environment (Supv)	.35	-.27	.17	.71	.16		.04	.01	.04	.01
35	FX Disciplined Initiative (Peer)	.10	-.17	.62	.24	.60	.19		.04	.03	.04
36	FX Disciplined Initiative (Supv)	.29	-.27	.22	.76	.19	.74	.23		.03	.01
37	FX Communication Method and Frequency (Peer)	.14	-.12	.55	.24	.60	.21	.62	.21		.04
38	FX Communication Method and Frequency (Supv)	.33	-.26	.23	.72	.17	.74	.24	.72	.21	

Note. Values below the diagonal are means of the correlations across 500 random datasets (n = 370). Values above the diagonal are standard deviations of the correlations across 500 datasets.

APPENDIX B

DERIVATIONS OF FORMULAS TO ESTIMATE PERFORMANCE CRITERION RELIABILITIES

Formula for the Composite Score

$$Y = \sum_{i=1}^{P}\left[w_i \left(\sum_{a=1}^{m} r_{ia} + \sum_{b=1}^{n} r_{ib} \right) \right] + \sum_{k=1}^{Q} z_k$$

(1)

where: Y = Composite score;
P = Number of rating dimensions for the composite (e.g., for the GTP composite, P=6);
Q = Number of non-rating scores for the composite (e.g., for the GTP composite, Q=2);
m = Number of Peer Raters;
n = Number of Supervisor Raters;
w_i = $1/(m+n)$ = Weight of rating dimension i;
r_{ia} = Rating (standardized) of peer a on rating dimension i;
r_{ib} = Rating (standardized) of supervisor b on rating dimension i;
z_k = Score (standardized) on (non-rating) scale k;

Equation (1) above can be more generally written as follows:

$$Y = \sum_{i}^{R} w_i^* z_i^* \qquad (2)$$

where $S = R + Q$, with $R = P(m+n)$ = total number of ratings in the composite (so S is the total number of components on the right side of equation (1) above);
w_i^* = Weight of the component score:

$w_i^* = 1/(m+n)$ if $i \le R$;

$w_i^* = 1$ if $i > R$;

z_i^* = Standardized rating or score:

$z_i^* = r_i$ if $i \le R$;

$z_i^* = z_i$ if $i > R$.

A rating on dimension i can be decomposed into three components:

$$r_i = t_i + h_i + e_i \qquad (3)$$

where: t_i = True score of rating on dimension i;
h_i = Halo of rating on dimension i;
e_i = Residual of rating on dimension i;

A non-rating *k* can be decomposed into two components:

$$z_k = t_k + e_k \tag{4}$$

where: t_k = True score;
e_k = Residual.

Observed Variance

$$Var(Y) = Var\left(\sum_i^S w_i^* z_i^*\right) = Var\left[\sum_i^S w_i^* (t_i + h_i + e_i)\right] \tag{5}$$

where $h_i = 0$ if $i > (m+n)P$. All other notations are the same as above.

Expanding the above equation and simplifying the result, we have the formula for variance of Y as follows:[65]

$$Var(Y) = \sum_i^S \sum_j^S w_i^* w_j^* Cov(t_i, t_j) + \sum_i^R \sum_j^R w_i^* w_j^* Cov(h_i, h_j) + \sum_i^S w_i^{*2} Var(e_i) \tag{6}$$

All notations are as in previous equations.

It can be seen from equation (6) that the variance of the observed composite score has three components: (a) variance due to true score, (b) variance due to halo, and (c) residual variance.

True Score Variance

From equation (6) above:

$$Var(T) = \sum_i^S \sum_j^S w_i^* w_j^* Cov(t_i, t_j) = \sum_i^S \sum_j^S w_i^* w_j^* r_{t_i t_j} SD_{t_i} SD_{t_j} \tag{7}$$

where: $r_{t_i t_j}$ = Correlation between true score of component *i* and true score of component *j*;
($r_{t_i t_j}$ dimensions were estimated by the SEM model).

Because all the ratings and scores are standardized, we have:

$$SD_t = SD_z \sqrt{r_{zz}} = \sqrt{r_{zz}} \tag{8}$$

where: r_{zz_i} = Reliability of component *i*
(r_{zz_i} for the rating dimensions were estimated by the SEM model, which is the square of the loading of the respective true score on the dimension rating)

---

[65] Simplification was done based on following assumptions/rules:
Cov(t,h) = Cov(t,e) = Cov(h,e) = 0;
$Cov(e_i, e_j) = 0$ when $i \neq j$;
For raters a and a', $Cov(h_a, h_{a'}) = 0$ when $a \neq a'$.

Replacing (8) into (7):

$$Var(T) = \sum_{i}^{S}\sum_{j}^{S} w_i^* w_j^* r_{t_i t_j} \sqrt{r_{zz_i} r_{zz_j}} \quad (9)$$

Halo Variance

From equation (6) above:

$$Var(H) = \sum_{i}^{R}\sum_{j}^{R} w_i^* w_j^* Cov(h_i, h_j) \quad (10)$$

Because $R = P(m+n)$, equation (10) can be decomposed as follows:

$$Var(H) = \sum_{a}^{m}\sum_{i}^{P}\sum_{j}^{P} w_i^* w_j^* Cov(h_{ia}, h_{ja}) + \sum_{b}^{n}\sum_{i}^{P}\sum_{j}^{P} w_i^* w_j^* Cov(h_{ib}, h_{jb}) \quad (11)$$

where the first component of the right side of equation (11) represents halo variance in peer ratings and the second component represents halo variance due to supervisor ratings.

Because it is assumed that halo is the same for all peers, we have:
$Cov(h_{ia}, h_{ja}) = Cov(h_{ia'}, h_{ja'})$ for all *a, a', i,* and *j.*

Similarly, it is assumed that halo is the same for all supervisors:
$Cov(h_{ib}, h_{jb}) = Cov(h_{ib'}, h_{jb'})$ for all *b, b', i*, and *j*.

Also, because this halo component only consists of rating dimensions:

$$w_i^* = w_j^* = \frac{1}{m+n} \text{ for all } w_i^*, w_j^*$$

Equation (10) can therefore be re-written as follows:

$$Var(H) = \frac{m\sum_{i}^{P}\sum_{j}^{P} Cov(h_{ia}, h_{ja}) + n\sum_{i}^{P}\sum_{j} Cov(h_{ib}, h_{jb})}{(m+n)^2} \quad (12)$$

Call the halo loading of peer rating dimension *i* estimated by the SEM model hl_{ia} and halo loading of peer rating on dimension *j* hl_{ja}, covariance due to halo between dimensions *i* and *j* is then:

$$Cov(h_{ia}, h_{ja}) = hl_{ia} hl_{ja} \quad (13)$$

Similarly, call the halo loading of supervisor rating dimension *i* estimated by the SEM model hl_{ib} and halo loading of supervisor rating on dimension *j* hl_{jb}, covariance due to halo between dimensions *i* and *j* is then:

$$Cov(h_{ib}, h_{jb}) = hl_{ib} hl_{jb} \quad (14)$$

Replacing equations (13) and (14) into equation (12) and simplifying, we have:

$$Var(H) = \frac{m(\sum_{i}^{P} hl_{ia})^2 + n(\sum_{i}^{P} hl_{ib})^2}{(m+n)^2} \quad (15)$$

Residual Variance

From equation (6) above:

$$Var(E) = \sum_{i}^{S} w_i^{*2} Var(e_i) \quad (16)$$

$Var(e_i)$ is estimated by the SEM model.

Reliability of the composite can be estimated by dividing the variance due to true score by the observed variance:

$$R_{yy} = Var(T) / Var(Y) = Var(T) / [Var(T)+Var(H)+Var(E)] \quad (17)$$

Var(T), Var(H), and Var(E) are estimated by equations (9), (15), and (16) above, respectively. [66]

Table B.1 shows reliability estimates for the performance composites in the Wave 1, Wave 2, and full samples. The values were obtained following the procedure described in this appendix.

Table B.1. Reliability Estimates for Performance Composites

	Reliability		
Performance Composite	Wave 1	Wave 2	Full Sample
General Technical Proficiency (GTP)	.708	.443	.685
Achievement and Effort (w/ CSJT)	.818	.793	.796
Achievement and Effort (w/o CSJT)	.785	.767	.770
Physical Fitness (PF)	.311	.397	.348
Teamwork (TM)	.930	.899	.920
Future Expected Performance (FXP)	.548	.324	.544

[66] All the equations were set up in an Excel spreadsheet for each composite to automate the calculations.

APPENDIX C

TABLES TO ACCOMPANY CHAPTER 13

Overview

This appendix provides correlations among (a) predictor scales and (b) predictor composites discussed in Chapter 13. Correlations were corrected for direct restriction of range (Thorndike's [1949] case 2) when AFQT was one of the variables correlated, and for indirect range restriction (Thorndike's case 3) when AFQT was not among the variables in the correlation.

The first four tables present the raw and corrected correlations between predictor scale scores, as shown in Figure C.1:

- Table C.1 provides correlations between the ASVAB, Target Tracking (TT), PSJT and all other predictor scale scores.
- Table C.2 provides correlations between the WSI and the RBI, WPS, and WVI scale scores.
- Table C.3 provides correlations between the RBI and the WPS and WVI scale scores.
- Table C.4 presents the correlations between the WPS and WVI scale scores.

The remaining three tables (Tables C.5-C.7) present correlations between predictor composite scores.

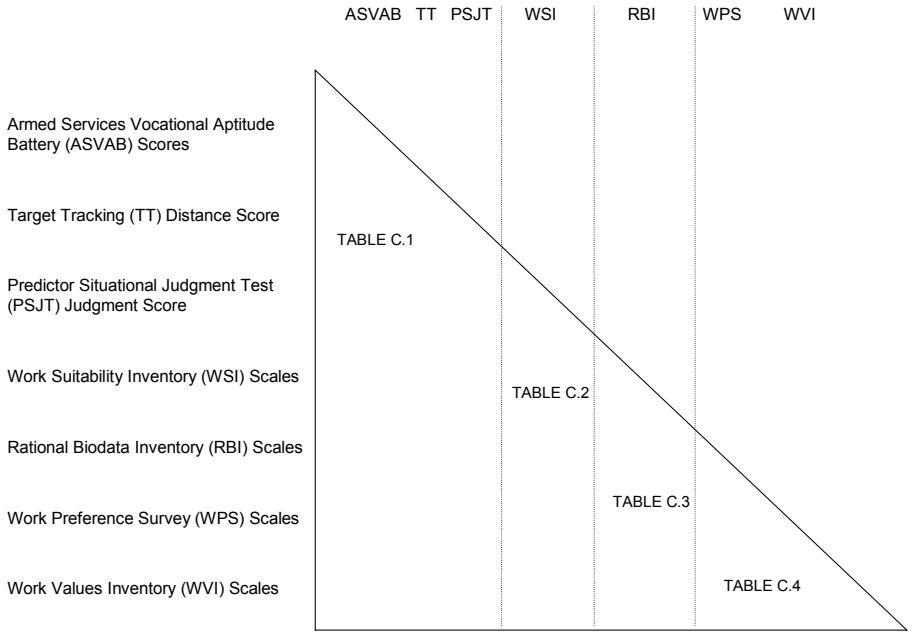


Figure C.1. Portions of the predictor scale score correlation matrix in Tables C.1-C.4.

Table C.1. Correlations between ASVAB, Target Tracking, and PSJT Judgment Scores and the WSI, RBI, WPS, and WVI Scale Scores

Score	ASVAB AFQT	ASVAB Spatial	ASVAB Technical	Target Tracking	PSJT Judgment
ASVAB Scores					
Spatial	**.38** (.55)				
Technical	**.51** (.68)	**.46** (.58)			
Target Tracking Distance	**.19** (.29)	**.30** (.36)	**.34** (.40)		
PSJT Judgment	**.22** (.34)	**.10** (.20)	**.09** (.21)	**.18** (.23)	
WSI					
Achievement/Effort	-.03 (-.05)	-.05 (-.06)	-.03 (-.04)	.01 (.00)	.03 (.02)
Adaptability/Flexibility	-.04 (-.06)	-.03 (-.04)	-.08 (-.09)	-.07 (-.08)	-.01 (-.02)
Attention to Detail	-.01 (-.01)	.00 (-.01)	.05 (.04)	.05 (.05)	.06 (.06)
Concern for Others	**-.15** (-.23)	-.09 (-.15)	**-.22** (-.28)	**-.12** (-.16)	.01 (-.04)
Cooperation	**-.17** (-.26)	**-.12** (-.19)	**-.20** (-.28)	**-.10** (-.14)	.01 (-.05)
Dependability	-.02 (-.02)	-.03 (-.04)	-.04 (-.05)	-.02 (-.02)	.04 (.03)
Energy	**-.08** (-.13)	-.04 (-.07)	-.02 (-.06)	.04 (.02)	-.07 (-.09)
Independence	**.16** (.25)	**.10** (.17)	**.20** (.26)	**.10** (.14)	-.05 (.00)
Initiative	.01 (.02)	.06 (.06)	.05 (.05)	-.05 (-.04)	.01 (.01)
Innovation	**.12** (.18)	**.15** (.19)	**.11** (.17)	-.01 (.02)	-.02 (.02)
Leadership Orientation	.02 (.03)	.07 (-.07)	.02 (-.03)	**.08** (.08)	.04 (.04)
Persistence	.06 (.10)	.01 (.04)	**.16** (.17)	.06 (.08)	**-.09** (-.07)
Self-Control	.07 (.11)	-.02 (.02)	.07 (.11)	.03 (.05)	.03 (.05)
Social Orientation	-.02 (-.03)	.05 (.04)	-.01 (-.02)	-.03 (-.04)	.04 (.04)
Stress Tolerance	**.09** (.15)	-.05 (.00)	**.12** (.16)	.04 (.07)	**-.11** (-.08)
Cultural Tolerance	-.01 (-.01)	-.01 (-.02)	**-.14** (-.12)	-.02 (-.02)	.07 (.06)
RBI (*lie adjusted*)					
Peer Leadership	**.16** (.24)	.08 (.15)	**.12** (.20)	**.11** (.15)	**.14** (.18)
Cognitive Flexibility	**.32** (.47)	**.20** (.32)	**.15** (.31)	**.16** (.22)	**.28** (.35)
Achievement	.08 (.12)	.03 (.07)	-.04 (.02)	.01 (.03)	**.27** (.28)
Fitness Motivation	.04 (.06)	.07 (.08)	.04 (.06)	**.12** (.13)	**.10** (.11)
Interpersonal Skills – Diplomacy	.05 (.07)	-.01 (.01)	-.02 (.01)	-.02 (.00)	**.17** (.17)
Stress Tolerance	**.14** (.22)	**.09** (.15)	**.10** (.17)	**.12** (.16)	**.11** (.15)
Hostility to Authority	**-.15** (-.24)	-.05 (-.12)	-.04 (-.13)	**-.11** (-.14)	**-.35** (-.38)
Self-Efficacy	**.15** (.23)	.06 (.13)	**.10** (.18)	**.09** (.12)	**.17 (** .21)
Cultural Tolerance	.05 (.07)	.01 (.04)	**-.11** (-.06)	.06 (.07)	**.30** (.31)
Internal Locus of Control	**.15** (.23)	.05 (.12)	.06 (.14)	**.15** (.19)	**.25** (.28)
Army Identification	.02 (.04)	.03 (.04)	**.10** (.10)	**.09** (.09)	**.16** (.16)
Respect for Authority	.02 (.03)	.03 (.04)	-.02 (.00)	.03 (.04)	**.18** (.18)
Narcissism	-.03 (-.05)	-.03 (-.04)	**-.12** (-.12)	-.06 (-.07)	-.04 (-.05)
Gratitude	**.12** (.18)	.03 (.09)	.01 (.08)	.02 (.05)	**.30** (.32)
Lie Scale	**-.12** (-.19)	**-.10** (-.16)	**-.09** (-.15)	-.07 (-.10)	-.05 (-.09)
WPS Scale/Facet					
Realistic Interests	**-.14** (-.21)	**.10** (.02)	**.23** (.10)	**.11** (.07)	.02 (-.02)
Mechanical Facet	**-.08** (-.13)	**.14** (.08)	**.33** (.22)	**.11** (.08)	.01 (-.01)
Physical Facet	**-.12** (-.19)	.04 (-.02)	.04 (-.04)	**.08** (.05)	.03 (-.01)
Investigative Interests	**.11** (.16)	**.10** (.14)	.07 (.13)	.07 (.09)	**.14** (.16)
Critical Thinking Facet	**.18** (.28)	**.13** (.21)	**.15** (.24)	**.10** (.15)	**.22** (.27)
Conduct Research Facet	.01 (.01)	.04 (.04)	-.03 (-.02)	.02 (.02)	.02 (.02)
Artistic Interests	-.03 (-.06)	.07 (.05)	-.02 (-.04)	.00 (-.01)	-.04 (-.05)
Artistic Activities Facet	-.07 (-.11)	.03 (-.01)	-.07 (-.10)	-.01 (-.03)	**-.09** (-.11)
Creativity Facet	.05 (.08)	**.13** (.14)	**.10** (.12)	.02 (.04)	**.08** (.10)

Table C.1. (Continued)

Score	ASVAB AFQT	ASVAB Spatial	ASVAB Technical	Target Tracking	PSJT Judgment
Social Interests	**-.10** (-.16)	-.04 (-.08)	**-.18** (-.21)	-.06 (-.08)	**.22** (.18)
Work with Others Facet	**-.14** (-.21)	.02 (-.05)	**-.12** (-.19)	-.03 (-.07)	**.16** (.10)
Help Others Facet	-.05 (-.07)	-.06 (-.08)	**-.17** (-.18)	-.07 (-.08)	**.19** (.17)
Enterprising Interests	-.07 (-.11)	-.07 (-.10)	**-.12** (-.15)	-.02 (-.04)	**.12** (.10)
Prestige Facet	-.01 (-.02)	-.03 (-.03)	.04 (-.04)	.00 (.00)	**.20** (.18)
Lead Others Facet	**-.13** (-.20)	-.07 (-.13)	**-.08** (-.15)	.00 (-.04)	**.16** (.11)
High Profile Facet	-.07 (-.11)	-.07 (-.10)	**-.18** (-.19)	-.06 (-.08)	-.04 (-.06)
Conventional Interests	**-.12** (-.19)	**-.09** (-.15)	**-.18** (-.23)	-.06 (-.09)	**.15** (.10)
Information Management Facet	-.07 (-.11)	-.07 (-.10)	**-.21** (-.22)	**-.08** (-.09)	.05 (.02)
Detail Orientation Facet	**-.10** (-.15)	-.02 (-.07)	.00 (-.07)	.01 (-.02)	**.20** (.16)
Clear Procedures Facet	**-.13** (.20)	**-.10** (-.15)	**-.08** (-.15)	-.03 (-.06)	**.20** (.15)
WVI Scales					
Social Status	-.01 (-.01)	.00 (.00)	-.03 (-.03)	**.09** (.09)	**.19** (.18)
Advancement	.06 (.10)	.04 (.07)	.02 (.06)	**.08** (.09)	**.21** (.22)
Autonomy	**.19** (.29)	**.12** (.21)	**.20** (.28)	**.12** (.16)	**.15** (.20)
Supportive Supervision	**-.09** (-.14)	.00 (-.04)	**-.11** (-.15)	.03 (.00)	**.17** (.14)
Leisure Time	**.17** (.27)	**.13** (.20)	**.16** (.24)	**.10** (.15)	**.09** (.14)
Comfort	.07 (.11)	**.10** (.12)	.02 (.06)	.04 (.06)	**.12** (.13)
Achievement	**.14** (.21)	.07 (.13)	**.08** (.15)	**.08** (.12)	**.20** (.24)
Societal Contribution	.06 (.09)	.01 (.04)	.00 (.03)	**.08** (.09)	**.22** (.23)
Independence	**.13** (.21)	.05 (.12)	**.16** (.22)	**.08 (** .11)	.01 (.05)
Social Service	-.02 (-.04)	-.01 (-.02)	.07 (-.08)	.05 (.04)	**.23** (.21)
Fixed Role	.05 (.09)	.01 (.04)	.05 (.08)	**.10** (.11)	**.15** (.16)
Variety	.07 (.11)	.06 (.09)	**.10** (.13)	**.08** (.10)	**.12** (.13)
Leadership Opportunities	.00 (.00)	.03 (.03)	.02 (.02)	.06 (.06)	**.22** (.21)
Feedback	.03 (.05)	.03 (.05)	.03 (.05)	.07 (.07)	**.16** (.16)
Travel	.00 (.00)	-.01 (-.01)	.03 (.03)	.02 (.02)	.06 (.06)
Physical Development	-.03 (-.05)	.02 (.00)	-.04 (-.06)	**.08** (.07)	**.12** (.10)
Ability Utilization	**.21** (.32)	**.16** (.25)	**.17** (.27)	**.15** (.19)	**.23** (.28)
Creativity	**.18** (.27)	**.19** (.25)	**.16** (.24)	**.08** (.12)	**.14** (.18)
Recognition	.04 (.06)	.05 (.07)	.02 (.04)	.03 (.04)	**.13** (.14)
Co-Workers	**.12** (.18)	**.12** (.17)	.07 (.13)	**.12** (.15)	**.18** (.21)
Activity	.06 (.09)	.03 (.06)	**.11** (.13)	**.08** (.09)	**.15** (.16)
Flexible Schedule	**.10** (.15)	.08 (.12)	.07 (.12)	**.10** (.12)	**.14** (.16)
Personal Development	.07 (.11)	**.13** (.15)	.07 (.11)	**.09** (.10)	**.19** (.21)
Home	**.18** (.27)	.09 (.17)	**.13** (.22)	**.12** (.16)	**.17** (.21)
Esteem	**.16** (.24)	.07 (.14)	.07 (.16)	**.08** (.12)	**.21** (.25)
Emotional Development	.06 (.10)	.03 (.06)	.00 (.04)	.06 (.08)	**.23** (.24)
Influence	**.09** (.14)	.06 (.10)	.06 (.11)	.05 (.07)	**.15** (.17)
Team Orientation	.04 (.07)	.08 (.09)	.01 (.03)	.06 (.07)	**.16** (.17)

Note. Bold indicates $p < .05$, two-tailed. $n = 470 – 755$ for raw correlations. Corrected correlations are in parentheses.

Table C.2. Correlations between WSI Scale Scores and RBI, WPS, and WVI Scale Scores

RBI, WPS, and WVI Scales	Work Suitability Inventory							
	Achievement/ Effort	Adaptability/ Flexibility	Attention to Detail	Concern for Others	Cooperation	Dependability	Energy	Independence
RBI (*lie adjusted*)								
Peer Leadership	-.02 (-.03)	**-.13** (-.14)	-.01 (-.01)	**-.16** (-.19)	**-.11** (-.14)	-.02 (-.03)	-.03 (-.05)	.01 (.05)
Cognitive Flexibility	.07 (.05)	.00 (-.02)	.01 (.01)	**-.10** (-.16)	**-.09** (-.15)	-.08 (-.08)	**-.12** (-.15)	.02 (.09)
Achievement	**.13** (.12)	-.08 (-.08)	**.13** (.13)	**-.11** (-.12)	-.01 (-.03)	.08 (.07)	.00 (.00)	**-.12** (-.10)
Fitness Motivation	**.14** (.14)	.00 (.00)	.06 (.06)	**-.28** (-.29)	-.02 (-.03)	.00 (.00)	**.17** (.17)	-.04 (-.03)
Interpersonal Skills – Diplomacy	-.03 (-.03)	.00 (.00)	-.06 (-.06)	-.04 (-.05)	.02 (.01)	**-.11** (-.11)	-.01 (-.01)	**-.18** (-.17)
Stress Tolerance	.03 (.03)	.05 (.04)	.02 (.02)	**-.14** (-.17)	-.05 (-.08)	-.05 (-.05)	.02 (.00)	.02 (.05)
Hostility to Authority	**-.11** (-.10)	.02 (.02)	**-.10 (**-.09)	-.05 (-.01)	-.06 (-.02)	-.05 (-.04)	.06 (.08)	.06 (.03)
Self-Efficacy	**.13** (.13)	-.08 (-.08)	**.13** (.12)	**-.22** (-.24)	**-.14** (-.17)	.02 (.02)	.04 (.03)	.01 (.04)
Cultural Tolerance	-.04 (-.04)	.05 (.05)	-.03 (-.03)	.04 (.03)	.02 (.01)	**-.10** (-.11)	-.04 -.05)	**-.15** (-.14)
Internal Locus of Control	**.13** (.12)	.00 (.00)	**.08** (.08)	-.03 (-.06)	-.04 (-.07)	.05 (.05)	-.02 (-.03)	-.06 (-.02)
Army Identification	.04 (.04)	-.06 (-.06)	.07 (.07)	**-.26** (-.26)	-.06 (-.06)	.07 (.07)	**.18** (.17)	-.08 (-.07)
Respect for Authority	.02 (.02)	-.03 (-.03)	**.08** (.08)	-.02 (-.02)	-.03 (-.04)	.06 (.06)	**.11** (.11)	**-.14** (-.13)
Narcissism	.07 (.07)	-.09 (-.08)	**.10** (.10)	**-.11** (-.10)	-.03 (-.02)	.02 (.03)	.07 (.08)	.00 (-.01)
Gratitude	.01 (.00)	-.03 (-.04)	.00 (.00)	-.04 (-.06)	.03 (.00)	.02 (.02)	.02 (.01)	**-.15** (-.12)
Lie Scale	.04 (.05)	.00 (.00)	.05 (.06)	-.03 (.00)	**.09** (.11)	.05 (.05)	-.02 (.00)	-.03 (-.05)
WPS Scale/Facet								
Realistic Interests Scale	.01 (.02)	-.01 (-.01)	.07 (.07)	**-.25** (-.21)	**-.11** (-.07)	-.03 (-.03)	**.27** (.28)	-.05 (-.08)
Mechanical Facet	-.02 (-.01)	-.05 (-.05)	**.10** (.10)	**-.18** (-.16)	**-.09** (-.07)	.00 (.00)	**.10** (.11)	.01 (-.01)
Physical Facet	.05 (.06)	.01 (.02)	.02 (.02)	**-.26** (-.22)	**-.10** (-.06)	-.04 (-.03)	**.36** (.36)	**-.09** (-.12)
Investigative Interests Scale	.02 (.01)	-.04 (-.04)	.06 (.06)	-.05 (-.07)	-.03 (-.06)	-.06 (-.06)	**-.11** (-.12)	-.01 (.02)
Critical Thinking Facet	.01 (.00)	-.07 (-.07)	**.10** (.09)	**-.13** (-.16)	**-.10** (-.14)	-.04 (-.04)	-.06 (-.08)	.02 (.06)
Conduct Research Facet	.02 (.02)	.00 (.00)	.01 (.01)	.04 (.04)	.04 (.03)	-.06 (-.06)	**-.13** (-.13)	-.03 (-.03)
Artistic Interests Scale	**-.08** (.08)	-.02 (-.02)	**-.13** (-.12)	**.12** (.13)	.02 (.03)	**-.15** (-.15)	**-.11** (-.11)	-.02 (-.03)
Artistic Activities Facet	-.05 (-.04)	-.01 (-.01)	**-.12** (-.12)	**.14** (.15)	.05 (.07)	**-.12** (-.12)	**-.11** (-.10)	-.02 (-.04)
Creativity Facet	**-.12** (-.12)	-.03 (-.03)	**-.08** (-.08)	.03 (.02)	-.04 (-.05)	**-.15** (-.15)	-.07 (-.07)	-.01 (.00)
Social Interests Scale	-.03 (-.02)	.01 (.01)	-.05 (-.05)	**.14** (.15)	**.11** (.13)	**-.09** (-.09)	-.07 (-.05)	**-.21** (-.23)
Work with Others Facet	-.06 (-.05)	.02 (.03)	-.02 (-.02)	.03 (.06)	.06 (.09)	**-.08** (-.08)	.02 (.03)	**-.25** (-.27)
Help Others Facet	-.02 (-.02)	.02 (.02)	**-.10** (-.10)	**.22** (.23)	**.10** (.11)	**-.10** (-.10)	**-.14** (-.13)	**-.13** (-.14)
Enterprising Interests Scale	.01 (.02)	-.06 (-.06)	.02 (.02)	-.06 (-.04)	.03 (.05)	-.03 (-.03)	-.04 (-.03)	-.05 (-.07)
Prestige Facet	.02 (.02)	**-.08** (-.08)	.04 (.04)	-.04 (-.03)	.01 (.01)	.04 (.04)	.02 (.02)	-.01 (-.01)
Lead Others Facet	.01 (.02)	**-.10** (-.09)	.02 (.02)	**-.09** (-.06)	.00 (.03)	.00 (.00)	.03 (.05)	**-.10** (-.13)
High Profile Facet	.03 (.04)	.04 (.04)	-.01 (-.01)	.00 (.01)	.07 (.09)	**-.09** (-.09)	**-.11** (-.10)	-.02 (-.04)

Table C.2. (Continued)

RBI, WPS, and WVI Scales	Work Suitability Inventory							
	Achievement/ Effort	Adaptability/ Flexibility	Attention to Detail	Concern for Others	Cooperation	Dependability	Energy	Independence
Conventional Interests Scale	.06 (.07)	-.05 (-.04)	**.20** (.20)	-.02 (.00)	**.08** (.11)	**.08** (-.08)	-.05 (-.04)	**-.09** (-.11)
Information Mgmt. Facet	.06 (.06)	.01 (.01)	**.10** (.10)	.05 (.06)	**.08** (.10)	.01 (.01)	**-.13** (-.12)	-.06 (-.07)
Detail Orientation Facet	.04 (.04)	**-.11** (-.10)	**.24** (.24)	**-.14** (-.11)	.01 (.03)	**.09** (.09)	.05 (.06)	-.05 (-.07)
Clear Procedures Facet	.02 (.03)	**-.09** (-.08)	**.21** (.21)	-.06 (-.03)	.04 (.07)	**.10** (.10)	.05 (.06)	**-.08** (-.10)
WVI Scales								
Social Status	.06 (.06)	-.02 (-.02)	**.10** (.10)	.06 (.06)	.05 (.05)	.01 (.01)	.01 (.01)	**-.09** (-.09)
Advancement	**.10** (.10)	-.01 (-.01)	**.10** (.10)	-.05 (-.06)	.00 (-.01)	.05 (.05)	.04 (.03)	-.04 (-.03)
Autonomy	-.01 (.07)	-.03 (-.04)	.05 (.05)	-.05 (-.08)	**-.08** (-.12)	-.04 (-.04)	-.06 (-.08)	**.17** (.20)
Supportive Supervision	.05 (.05)	.00 (.00)	.07 (.07)	.04 (.06)	**.08** (.10)	-.02 (-.02)	.05 (.06)	**-.15** (-.17)
Leisure Time	-.04 (-.04)	-.01 (-.02)	-.01 (-.02)	.02 (-.02)	.01 (-.04)	-.02 (-.03)	-.04 (-.06)	.07 (.11)
Comfort	.01 (.01)	.03 (.03)	.02 (.02)	**.12** (.10)	**.09** (.07)	-.04 (-.04)	**-.14** (-.14)	.06 (.08)
Achievement	**.11** (.10)	-.01 (-.02)	**.09** (.08)	-.02 (-.05)	-.02 (-.05)	.04 (.03)	.02 .00)	-.03 (-.00)
Societal Contribution	.07 (.07)	-.03 (-.04)	.06 (.06)	.05 (.03)	.02 (.00)	.01 (.01)	.00 (-.01)	**-.09** (-.07)
Independence	-.02 (-.03)	-.02 (-.02)	**.09** (.08)	-.07 (-.10)	-.04 (-.08)	-.01 (-.01)	-.02 (-.04)	**.26** (.28)
Social Service	.07 (.07)	-.02 (-.02)	**.09** (.09)	**.17** (.17)	.05 (.06)	-.01 (-.01)	-.07 (-.06)	**-.15** (-.15)
Fixed Role	.05 (.04)	-.07 (-.07)	**.16** (.16)	-.07 (-.08)	-.03 (-.04)	.04 (.04)	.05 (.05)	.00 (.02)
Variety	-.02 (-.02)	.06 (.06)	.05 (.05)	-.04 (-.06)	-.02 (-.03)	-.07 (-.07)	.03 (.02)	-.02 (.00)
Leadership Opportunities	**.08** (.08)	-.06 (-.06)	.08 (.08)	**-.14** (-.14)	-.07 (-.07)	.04 (.04)	.04 (.04)	-.04 (-.04)
Feedback	.07 (.07)	-.04 (-.04)	**.14** (.14)	-.02 (-.03)	-.01 (-.02)	.03 (.03)	-.02 (-.03)	-.02 (-.01)
Travel	.00 (.00)	.01 (.01)	.01 (.01)	**-.09** (-.09)	-.01 (-.01)	-.06 (-.06)	.02 (.02)	-.01 (-.01)
Physical Development	**.08** (.08)	-.01 (-.01)	.06 (.07)	**-.11** (-.10)	-.01 (-.01)	-.02 (-.02)	**.26** (.26)	**-.11** (-.11)
Ability Utilization	**.08** (.07)	-.05 (-.06)	**.17** (.16)	**-.09** (-.13)	-.07 (-.12)	.00 (.00)	-.03 (-.05)	.03 (.07)
Creativity	.05 (.04)	-.01 (-.02)	.06 (.05)	-.04 (-.07)	-.06 (-.10)	-.08 (-.08)	**-.10** (-.12)	**.11** (.15)
Recognition	.08 (.07)	.00 (.00)	**.11** (.11)	.04 (.03)	**.10** (.09)	.05 (.05)	-.02 (-.02)	-.01 (.00)
Co-Workers	.01 (.01)	-.01 (-.02)	.04 (.03)	.08 (.05)	**.09** (.06)	.00 (.00)	-.03 (-.04)	-.06 (-.03)
Activity	**.13** (.13)	-.03 (-.03)	**.15** (.15)	-.05 (-.06)	-.05 (-.07)	.07 (.07)	.02 (.01)	.04 (.05)
Flexible Schedule	-.02 (-.02)	.04 (.03)	.01 (.01)	.02 (-.01)	.03 (.00)	-.07 (-.07)	**-.08** (-.09)	.03 (.06)
Personal Development	.07 (.07)	-.02 (-.02)	**.16** (.16)	-.04 (-.06)	-.03 (-.05)	.04 (.04)	-.06 (-.06)	-.01 (.01)
Home	.03 (.02)	-.05 (-.06)	.07 (.06)	.05 (.01)	.05 (.00)	.02 (.02)	-.04 (-.06)	.03 (.07)
Esteem	.04 (.03)	-.02 (-.03)	**.12** (.12)	.04 (.01)	.02 (-.02)	-.01 (-.01)	-.05 (-.07)	-.03 (.00)
Emotional Development	**.10** (.09)	-.07 (-.07)	**.17** (.16)	-.07 (-.08)	-.02 (-.04)	.07 (.07)	.02 (.02)	**-.13** (-.16)
Influence	.06 (.06)	-.04 (-.04)	**.14** (.14)	-.05 (-.07)	-.05 (-.07)	.05 (.05)	.00 (-.01)	-.04 (-.01)
Team Orientation	.04 (.04)	.03 (.03)	.05 (.05)	**.10** (.09)	**.09** (.07)	-.02 (-.02)	-.06 (-.07)	**-.17** (-.16)

Table C.2. (Continued)

RBI, WPS, and WVI Scales	Work Suitability Inventory (Continued)							
	Initiative	Innovation	Leadership Orientation	Persistence	Self-Control	Social Orientation	Stress Tolerance	Cultural Tolerance
RBI (*lie adjusted*)								
Peer Leadership	.06 (. 06)	.07 (. 09)	**.27** (. 27)	.01 (. 02)	.06 (. 08)	-.03 (-.04)	**.11** (.13)	-.03 (-.03)
Cognitive Flexibility	-.02 (-.02)	**.18** (. 22)	**.09** (. 09)	.02 (. 05)	.02 (. 05)	**-.09** (-.09)	.05 (.09)	.04 (. 03)
Achievement	.06 (. 06)	-.07 (-.06)	**.15** (. 16)	-.02 (-.02)	-.06 (-.05)	-.02 (-.02)	.05 (.06)	-.06 (-.06)
Fitness Motivation	.03 (. 03)	**-.14** (-.13)	**.14** (. 14)	-.02 (-. 01)	.00 (. 00)	-.03 (-.03)	**.12** (.13)	**-.09** (-.09)
Interpersonal Skills – Diplomacy	.00 (. 00)	.05 (. 05)	**.15** (. 15)	-.06 (-. 06)	-.04 (-.03)	**.15** (. 15)	.02 (.03)	**.12** (. 12)
Stress Tolerance	.02 (. 02)	-.04 (-.01)	-.01 (-.01)	-.02 (-.01)	.07 (. 09)	-.02 (-.02)	.08 (.10)	.03 (. 03)
Hostility to Authority	.06 (. 05)	.06 (. 03)	**.13** (. 12)	**.09** (. 08)	.02 (. 00)	-.04 (-.03)	.04 (.02)	**-.09** (-.09)
Self-Efficacy	.06 (. 07)	-.01 (. 02)	**.16** (. 16)	-.03 (-.01)	.03 (. 05)	**-.08** (-.09)	**.11** (.13)	**-.10** (-.10)
Cultural Tolerance	-.02 (- 01)	-.02 (-.01)	.01 (. 01)	**-.11** (-.10)	-.02 (-.01)	.04 (. 04)	-.02 (-.02)	**.34** (. 34)
Internal Locus of Control	.04 (.05)	**-.12** (-.09)	.05 (. 05)	-.08 (-.06)	.00 (. 01)	-.02 (-.02)	.02 (. 04)	.00 (. 00)
Army Identification	**.11** (.11)	**-.13** (-.13)	**.16** (. 16)	-.04 (-.03)	.03 (. 03)	.00 (. 00)	**.18** (. 18)	**-.13** (-.13)
Respect for Authority	.05 (.05)	-.08 (-.07)	**.10** (. 10)	-.02 (-.02)	-.04 (-.04)	.03 (. 03)	-.04 (-.04)	-.03 (-.03)
Narcissism	.00 (.00)	.01 (. 01)	**.17** (. 17)	-.05 (-.05)	-.06 (-.06)	-.05 (-.05)	.03 (. 02)	-.07 (-.07)
Gratitude	**.08** (.08)	-.03 (-.01)	.04 (. 04)	**-.09** (-.08)	.01 (. 02)	**.09** (.09)	-.01 (. 00)	.07 (. 07)
Lie Scale	.00 (.00)	-.08 (-.09)	-.02 (-.02)	.02 (.01)	.01 (. 00)	-.04 (-.04)	.01 (-.01)	-.05 (-.05)
WPS Scale/Facet								
Realistic Interests Scale	**.10** (.10)	-.05 (-.07)	.06 (.06)	**.12** (.11)	.01 (-.01)	-.02 (-.02)	**.12** (.10)	**-.16** (-.16)
Mechanical Facet	.05 (.05)	.02 (.01)	.01 (.01)	**.17** (.16)	.01 (.00)	.00 (.00)	.05 (.04)	**-.13** (-.13)
Physical Facet	**.12** (.12)	**-.13** (-.15)	**.10** (.10)	.04 (.03)	.03 (.02)	-.04 (-.04)	**.15** (.13)	**-.15** (-.14)
Investigative Interests Scale	-.01 (-.01)	.05 (.07)	**.10** (.10)	.07 (.08)	.03 (.04)	-.05 (-.05)	.02 (.03)	.03 (.03)
Critical Thinking Facet	.03 (.03)	.05 (.07)	**.10** (.10)	**.09** (.10)	.05 (.07)	**-.09** (-.09)	**.08** (.10)	.00 (-.01)
Conduct Research Facet	-.05 (-.05)	.05 (.05)	.06 (.06)	.04 (.04)	-.01 (-.01)	.00 (.00)	-.04 (-.04)	.05 (.05)
Artistic Interests Scale	-.05 (-.05)	**.30** (.29)	**.08** (.08)	-.02 (-.02)	-.04 (-.04)	.05 (.05)	**-.10** (-.10)	**.10** (.10)
Artistic Activities Facet	-.07 (-.07)	**.23** (.21)	.05 (.05)	-.03 (-.03)	-.02 (-.03)	.05 (.05)	**-.11** (-.12)	**.10** (.10)
Creativity Facet	.01 (.01)	**.32** (.32)	**.10** (.10)	.00 (.01)	-.05 (-.04)	.04 (.04)	-.03 (-.03)	.06 (.06)
Social Interests Scale	-.04 (-.04)	-.03 (-.04)	**.11** (.10)	**-.10** (-.11)	-.01 (-.02)	**.13** (.14)	**-.08** (-.09)	**.17** (.17)
Work with Others Facet	-.01 (-.01)	-.03 (-.06)	**.10** (.10)	-.06 (-.07)	-.03 (-.04)	**.16** (.16)	.00 (-.02)	**.14** (.14)
Help Others Facet	-.07 (-.07)	.00 (-.01)	.06 (.06)	**-.10** (-.11)	.01 (.01)	**.11** (.11)	**-.13** (-.13)	**.20** (.20)
Enterprising Interests Scale	.04 (.04)	.00 (-.01)	**.21** (.21)	-.04 (-.04)	-.02 (-.03)	.01 (.01)	.02 (.01)	-.02 (-.02)
Prestige Facet	.03 (.03)	.01 (.01)	**.14** (.14)	-.03 (-.04)	-.04 (-.04)	-.01 (-.01)	.00 (.00)	-.08 (-.08)
Lead Others Facet	**.11** (.11)	-.07 (-.09)	**.27** (.26)	-.06 (-.07)	-.01 (-.02)	.04 (.04)	.02 (.01)	-.03 (-.03)
High Profile Facet	-.05 (-.05)	.03 (.02)	**.09** (.09)	-.02 (-.03)	-.02 (-.02)	.01 (.02)	-.01 (-.02)	.04 (.04)

Table C.2. (Continued)

RBI, WPS, and WVI Scales	Work Suitability Inventory (Continued)							
	Initiative	Innovation	Leadership Orientation	Persistence	Self-Control	Social Orientation	Stress Tolerance	Cultural Tolerance
Conventional Interests Scale	-.04 (-.04)	**-.11** (-.13)	.03 (.03)	.00 (-.01)	-.04 (-.06)	.00 (.01)	-.05 (-.06)	.01 (.01)
Information Management Facet	**-.09** (-.09)	-.05 (-.06)	.02 (.02)	-.02 (-.03)	-.04 (-.04)	.06 (.06)	**-.09** (-.10)	.07 (.07)
Detail Orientation Facet	.03 (.03)	-.08 (-.09)	.03 (.03)	**.09** (.08)	-.04 (-.05)	**-.08** (-.07)	.05 (.03)	-.07 (-.07)
Clear Procedures Facet	.01 (.01)	**-.09** (-.11)	.03 (.03)	.02 (.01)	-.04 (-.05)	-.08 (-.07)	.00 (-.02)	-.02 (-.02)
WVI Scales								
Social Status	-.03 (-.03)	**-.08** (-.08)	.06 (.06)	-.04 (-.04)	.00 (.00)	.02 (.02)	-.06 (-.06)	-.04 (-.04)
Advancement	.00 (.00)	-.10 (-.09)	.07 (.07)	-.05 (-.05)	.01 (.02)	.00 (.00)	-.05 (-.05)	-.04 (-.04)
Autonomy	.02 (.02)	.06 (.09)	.01 (.01)	.01 (.02)	.05 (.07)	-.06 (-.06)	.00 (.03)	-.04 (-.04)
Supportive Supervision	.00 (.00)	-.06 (-.07)	.00 (.00)	-.06 (-.07)	-.02 (-.02)	**.13** (.13)	**-.12** (-.13)	-.01 (-.01)
Leisure Time	**-.09** (-.08)	**.08** (.11)	-.04 (-.03)	-.04 (-.03)	.07 (.09)	.04 (.03)	-.03 (-.01)	.01 (.01)
Comfort	**-.14** (-.14)	.06 (.07)	**-.12** (-.12)	-.06 (-.05)	.04 (.05)	**.11** (.11)	**-.16** (-.15)	.05 (.05)
Achievement	-.02 (-.02)	-.01 (.01)	.00 (.00)	-.02 (.00)	-.03 (-.01)	.00 (.00	-.03 (-.01)	-.05 (-.06)
Societal Contribution	-.02 (-.02)	-.03 (-.02)	.00 (.00)	-.07 (-.06)	.00 (.01)	-.01 (-.01)	-.07 (-.06)	**.10** (.10)
Independence	.00 (.00)	**.09** (.11)	-.03 (-.03)	-.01 (.00)	.00 (.02)	**-.13** (-.13)	-.01 (.01)	-.07 (-.07)
Social Service	-.06 (-.06)	-.07 (-.07)	-.02 (-.02)	-.08 (-.08)	-.02 (-.02)	**.08** (.08)	**-.12** (-.12)	**.11** (.11)
Fixed Role	.00 (.00)	-.07 (-.06)	.03 (.03)	-.02 (-.01)	.04 (.05)	.03 (.03)	-.06 (-.05)	**-.09** (-.09)
Variety	.01 (.02)	.05 (.06)	-.01 (-.01)	-.05 (-.04)	.04 (.04)	.02 (.02)	-.03 (-.02)	.01 (.01)
Leadership Opportunities	.06 (.06)	**-.10** (-.10)	**.28** (.28)	**-.09** (-.09)	.01 (.01)	.03 (.03)	.02 (.02)	**-.10** (-.10)
Feedback	-.03 (-.03)	.00 (.01)	.01 (.01)	-.06 (-.05)	.00 (.00)	.01 (.01)	-.06 (-.06)	.00 (.00)
Travel	.04 (.04)	.04 (.03)	-.02 (-.02)	-.01 (-.01)	.07 (.06)	-.04 (-.04)	.07 (.06)	.01 (.01)
Physical Development	.02 (.02)	**-.14** (-.15)	.05 (.04)	-.06 (-.07)	.00 (.00)	.00 (.00)	.03 (.02)	.00 (.00)
Ability Utilization	-.06 (-.05)	.05 (.08)	-.01 (.00)	.00 (.02)	.04 (.06)	.01 (.00)	-.07 (-.04)	.00 (.00)
Creativity	-.02 (-.02)	**.22** (.24)	.01 (.01)	-.03 (-.01)	-.01 (.01)	-.02 (-.03)	**-.09** (-.06)	-.01 (-.01)
Recognition	-.07 (-.07)	-.04 (-.03)	.00 (.00)	-.07 (-.06)	-.07 (-.07)	.02 (.02)	-.05 (-.05)	**-.09** (-.09)
Co-Workers	-.06 (-.06)	**-.09** (-.07)	-.02 (-.02)	**-.10** (-.09)	-.02 (-.01)	**.16** (.15)	**-.10** (-.08)	**.08** (.08)
Activity	.01 (.01)	-.07 (-.06)	-.06 (-.06)	.03 (.03)	-.01 (.00)	-.05 (-.05)	-.03 (-.02)	**-.09** (-.09)
Flexible Schedule	-.05 (-.05)	.05 (.06)	-.05 (-.05)	-.03 (-.02)	.05 (.06)	.07 (.07)	**-.09** (-.07)	.07 (.07)
Personal Development	-.03 (-.03)	.00 (.01)	-.01 (-.01)	-.02 (-.01	-.01 (.00)	.01 (.01)	-.08 (-.07)	.02 (.02)
Home	-.05 (-.05)	-.06 (-.03)	-.05 (-.04)	-.01 (.01)	.02 (.04)	.04 (.03)	-.07 (-.05)	.00 (.00)
Esteem	-.07 (-.07)	-.01 (.02)	.00 (.00)	-.05 (-.03)	.03 (.04)	.02 (.02)	-.05 (-.02)	.00 (.00)
Emotional Development	.03 (.03)	**-.11** (-.10)	.00 (.00)	-.05 (-.05)	.07 (.08)	.00 (.00)	.04 (.05)	-.02 (-.02)
Influence	.02 (.02)	-.04 (-.02)	**.09** (.09)	-.05 (-.04)	.05 (.06)	-.03 (-.03)	-.01 (.00)	**-.08** (-.08)
Team Orientation	-.03 (-.03)	-.08 (-.07)	-.05 (-.05)	**-.12** (-.11)	.04 (.04)	**.20** (.20)	-.06 (-.05)	.03 (.03)

Note. Bold indicates $p < .05$, two-tailed. $n = 487 – 658$ for raw correlations. Corrected correlations are in parentheses.

Table C.3. Correlations between RBI Scale Scores and WPS and WVI Scale Scores

	Rational Biodata Inventory Scales							
WPS and WVI Scales	Peer Leadership	Cognitive Flexibility	Achievement	Fitness Motivation	Int. Skills.-Diplomacy	Stress Tolerance	Hostility to Authority	Self-Efficacy
WPS Scale/Facet								
Realistic Interests Scale	.03 (.00)	.01 (-.05)	**.13** (.11)	.28 (.27)	.04 (.03)	.02 (-.01)	**.08** (.11)	**.11** (.08)
Mechanical Facet	-.02 (-.04)	-.02 (-.06)	.06 (.05)	**.11** (.11)	-.01 (-.02)	.01 (-.01)	.06 (.07)	.04 (.02)
Physical Facet	.07 (.04)	.06 (.00)	**.19** (.17)	**.39** (.38)	**.09** (.08)	.04 (.01)	.07 (.09)	**.16** (.13)
Investigative Interests Scale	**.33** (.35)	**.55** (.55)	**.37** (.37)	**.14** (.15)	**.16** (.17)	-.02 (.00)	**-.08** (-.10)	**.22** (.24)
Critical Thinking Facet	**.39** (.42)	**.55** (.58)	**.41** (.42)	**.19** (.19)	**.23** (.24)	.02 (.06)	-.08 (-.11)	**.32** (.34)
Conduct Research Facet	**.18** (.18)	**.40** (.38)	**.23** (.23)	.06 (.06)	.05 (.06)	-.06 (-.06)	-.06 (-.06)	.07 (.07)
Artistic Interests Scale	**.20** (.19)	**.33** (.29)	**.13** (.12)	-.04 (-.05)	**.11** (.10)	**-.12** (-.12)	**.08** (.08)	-.01 (-.02)
Artistic Activities Facet	.07 (.06)	**.19** (.15)	.03 (.02)	-.08 (-.09)	.01 (.00)	**-.13** (-.14)	**.09** (.11)	**-.12** (-.13)
Creativity Facet	**.36** (.36)	**.46** (.45)	**.27** (.27)	**.06** (.07)	**.26** (.26)	-.04 (-.03)	.00 (-.01)	**.21** (.22)
Social Interests Scale	**.30** (.27)	**.37** (.30)	**.41** (.39)	.16 (.15)	**.42** (.41)	.00 (-.02)	**-.12** (-.09)	**.23** (.21)
Work with Others Facet	**.24** (.20)	**.25** (.17)	**.31** (.29)	**.21** (.20)	**.46** (.44)	**.10** (.06)	-.04 (-.01)	**.23** (.19)
Help Others Facet	**.24** (.23)	**.33** (.29)	**.34** (.33)	**.04** (.04)	**.29** (.28)	-.07 (-.08)	**-.17** (-.16)	**.14** (.12)
Enterprising Interests Scale	**.36** (.33)	**.33** (.28)	**.38** (.37)	.18 (.18)	**.29** (.28)	-.07 (-.08)	.02 (.03)	**.29** (.27)
Prestige Facet	**.31** (.30)	**.28** (.26)	**.34** (.34)	**.16** (.16)	**.28** (.27)	-.07 (-.07)	-.03 (-.03)	**.31** (.30)
Lead Others Facet	**.33** (.29)	**.29** (.21)	**.41** (.39)	**.23** (.22)	**.33** (.32)	.01 (-.01)	-.04 (-.01)	**.33** (.29)
High Profile Facet	**.14** (.12)	**.15** (.11)	**.12** (.11)	**.04** (.04)	.06 (.06)	**-.08** (-.09)	**.09** (.10)	.05 (.03)
Conventional Interests Scale	**.17** (.14)	**.23** (.16)	**.39** (.37)	.14 (.13)	**.10** (.09)	-.07 (-.09)	**-.12** (-.09)	**.21** (.17)
Information Management Facet	**.08** (.07)	**.14** (.10)	**.24** (.23)	**.03** (.03)	.03 (.03)	-.07 (-.08)	-.05 (-.04)	.07 (.06)
Detail Orientation Facet	**.24** (.21)	**.31** (.24)	**.38** (.36)	.22 (.22)	**.19** (.18)	.00 (-.02)	**-.13** (-.10)	**.31** (.29)
Clear Procedures Facet	**.19** (.16)	**.25** (.17)	**.38** (.36)	**.21** (.20)	**.16** (.15)	-.03 (-.06)	**-.13** (-.10)	**.30** (.27)

Table C.3. (Continued)

WPS and WVI Scales	Rational Biodata Inventory Scales							
	Peer Leadership	Cognitive Flexibility	Achievement	Fitness Motivation	Int. Skills.-Diplomacy	Stress Tolerance	Hostility to Authority	Self-Efficacy
WVI Scales								
Social Status	.04 (.04)	.06 (.05)	**.14** (.14)	**.08** (.08)	**.12** (.11)	**-.08** (-.09)	-.02 (-.02)	.06 (.06)
Advancement	**.10** (.11)	**.08** (.10)	**.19** (.20)	**.14** (.14)	**.17** (.17)	-.03 (-.02)	**-.09** (-.10)	**.18** (.19)
Autonomy	.05 (.09)	.05 (.13)	.01 (.03)	**.08** (.09)	-.01 (.00)	.01 (.05)	-.06 (-.10)	.01 (.05)
Supportive Supervision	-.07 (-.09)	-.03 (-.07)	**.08** (.07)	**.01** (.00)	.07 (.06)	-.07 (-.09)	.01 (.03)	-.03 (-.05)
Leisure Time	-.06 (-.02)	-.03 (.05)	**-.14** (-.12)	.01 (.02)	.01 (.02)	-.06 (-.02)	.03 (-.03)	-.04 (.00)
Comfort	**-.11** (-.09)	-.05 (-.01)	**-.10** (-.09)	-.09 (-.08)	-.02 (-.01)	-.07 (-.05)	-.02 (-.05)	**-.10** (-.09)
Achievement	.07 (.10)	**.15** (.20)	**.16** (.18)	**.03** (.04)	.08 (.09)	-.06 (-.03)	-.07 (-.10)	.07 (.09)
Societal Contribution	.03 (.04)	**.15** (.17)	**.17** (.18)	.04 (.04)	.07 (.07)	-.03 (-.02)	**-.12** (-.13)	.03 (.04)
Independence	-.07 (-.04)	-.04 (.02)	**-.16** (-.14)	-.08 (-.07)	**-.19** (-.18)	-.07 (-.04)	.02 (-.01)	-.07 (-.04)
Social Service	.02 (.02)	**.11** (.09)	**.18** (.17)	**.00** (.00)	**.11** (.11)	-.05 (-.06)	**-.14** (-.13)	.03 (.03)
Fixed Role	-.01 (.00)	.01 (.04)	**.08** (.09)	.05 (.05)	.00 (.00)	-.07 (-.06)	-.06 (-.07)	.01 (.02)
Variety	-.02 (.00)	.06 (.08)	.01 (.02)	.06 (.07)	.03 (.04)	.02 (.03)	-.02 (-.03)	.02 (.03)
Leadership Opportunities	**.18** (.17)	**.12** (.11)	**.26** (.26)	.17 (.17)	**.19** (.19)	-.02 (-.02)	-.03 (-.02)	**.18** (.18)
Feedback	.03 (.04)	**.08** (.09)	**.13** (.13)	**.01** (.01)	.05 (.05)	-.06 (-.05)	**-.08** (-.09)	.05 (.05)
Travel	-.01 (-.01)	.07 (.06)	.00 (.00)	.03 (.03)	**.09** (.09)	.07 (.06)	.00 (.01)	.04 (.04)
Physical Development	-.03 (-.03)	-.02 (-.03)	**.10** (.09)	.33 (.32)	**.08** (.08)	.06 (.06)	-.01 (.00)	.07 (.06)
Ability Utilization	.05 (.09)	**.16** (.24)	.08 (.10)	**.06** (.07)	.03 (.05)	.05 (.09)	**-.11** (-.15)	.08 (.12)
Creativity	**.11** (.15)	**.18** (.24)	.01 (.03)	.01 (.02)	.02 (.03)	-.02 (.01)	.05 (.01)	.08 (.11)
Recognition	.03 (.04)	.04 (.05)	**.09** (.09)	-.01 (-.01)	.07 (.07)	**-.08** (-.07)	-.03 (-.04)	.02 (.03)
Co-Workers	.01 (.04)	.04 (.09)	.06 (.07)	.06 (.06)	**.11** (.12)	.00 (.02)	-.06 (-.09)	.05 (.07)
Activity	-.06 (-.04)	.03 (.05)	.01 (.02)	-.02 (-.02)	-.08 (-.07)	-.01 (.00)	**-.13** (-.14)	.01 (.02)
Flexible Schedule	**-.16** (-.13)	**-.09** (-.04)	**-.13** (-.12)	-.05 (-.04)	**-.08** (-.07)	.00 (.02)	-.03 (-.05)	**-.11** (-.09)
Personal Development	.00 (.02)	**.12** (.14)	**.11** (.11)	.05 (.05)	-.01 (-.01)	.04 (.06)	**-.08** (-.10)	**.09** (.10)
Home	-.01 (.03)	.02 (.10)	-.03 (-.01)	-.01(.00)	-.04 (-.03)	-.02 (.01)	-.07 (-.11)	-.03 (.01)
Esteem	.06 (.09)	**.09** (.15)	**.10** (.12)	-.01 (-.00)	.06 (.07)	.01 (.04)	-.07 (-.10)	.04 (.07)
Emotional Development	.00 (.01)	.05 (.07)	**.09** (.10)	.09 (.10)	.02 (.02)	.06 (.07)	**-.15** (-.16)	**.08** (.09)
Influence	.05 (.07)	.04 (.08)	.07 (.08)	**.03** (.03)	.03 (.04)	-.02 (.00)	-.03 (-.15)	.05 (.06)
Team Orientation	.03 (.04)	.05 (.06)	.08 (.08)	.06 (.06)	**.15** (.15)	.07 (.08)	**-.10** (-.11)	.03 (.03)

Table C.3. (Continued)

WPS and WVI Scales	Rational Biodata Inventory Scales (Continued)						
	Cultural Tolerance	Internal Locus of Control	Army Identification	Respect Authority	Narcissism	Gratitude	Lie Scale
WPS Scale/Facet							
Realistic Interests Scale	-.06 (-.07)	.04 (.01)	**.31** (.31)	**.18** (.17)	.02 (.03)	**.12** (.09)	.04 (.07)
Mechanical Facet	**-.08** (-.09)	.02 (.00)	**.15** (.14)	**.10** (.09)	-.03 (-.02)	.04 (.02)	.03 (.05)
Physical Facet	.00 (-.01)	.07 (.05)	**.39** (.38)	**.21** (.20)	**.08** (.08)	**.16** (.14)	.03 (.05)
Investigative Interests Scale	**.25** (.25)	**.12** (.14)	**.09** (.09)	**.22** (.22)	**.14** (.13)	.05 (.07)	.04 (.02)
Critical Thinking Facet	**.25** (.26)	**.18** (.21)	**.17** (.17)	**.23** (.23)	**.14** (.13)	**.11** (.14)	.03 (.00)
Conduct Research Facet	**.17** (.17)	.03 (.03)	-.01 (-.01)	**.15** (.15)	**.10** (.10)	-.02 (-.02)	.03 (.03)
Artistic Interests Scale	**.15** (.15)	-.05 (-.06)	-.06 (-.06)	.06 (.06)	**.11** (.12)	.03 (.02)	.00 (.00)
Artistic Activities Facet	.08 (.07)	**-.11** (-.12)	**-.09** (-.09)	.01 (.00)	.06 (.06)	-.02 (-.03)	-.02 (-.01)
Creativity Facet	**.24** (.24)	**.09** (.10)	.03 (.03)	**.15** (.15)	**.18** (.18)	**.12** (.13)	.03 (.02)
Social Interests Scale	**.37** (.36)	**.20** (.18)	**.17** (.17)	**.32** (.31)	**.15** (.15)	**.29** (.27)	.06 (.07)
Work with Others Facet	**.31** (.30)	**.21** (.17)	**.20** (.20)	**.27** (.26)	**.12** (.13)	**.30** (.27)	**.08** (.10)
Help Others Facet	**.32** (.31)	**.14** (.13)	**.09** (.09)	**.25** (.25)	**.10** (.10)	**.21** (.20)	.03 (.03)
Enterprising Interests Scale	**.20** (.19)	**.10** (.09)	**.15** (.14)	**.22** (.22)	**.36** (.36)	**.10** (.09)	.01 (.02)
Prestige Facet	**.18** (.18)	**.16** (.15)	**.17** (.16)	**.22** (.22)	**.32** (.32)	**.14** (.13)	.02 (.02)
Lead Others Facet	**.21** (.20)	**.19** (.15)	**.26** (.25)	**.26** (.26)	**.27** (.27)	**.17** (.15)	.04 (.07)
High Profile Facet	**.09** (.08)	-.06 (-.07)	-.06 (-.06)	.04 (.04)	**.22** (.22)	-.06 (-.07)	-.01 (.00)
Conventional Interests Scale	**.19** (.18)	**.08** (.05)	**.11** (.10)	**.25** (.25)	**.17** (.18)	**.10** (.08)	.07 (.09)
Information Management Facet	**.14** (.13)	.01 (.00)	-.05 (-.05)	**.13** (.13)	**.12** (.13)	.01 (.00)	.02 (.03)
Detail Orientation Facet	**.20** (.19)	**.19** (.16)	**.19** (.19)	**.27** (.26)	**.14** (.14)	**.14** (.12)	**.10** (.12)
Clear Procedures Facet	**.21** (.20)	**.15** (.12)	**.19** (.19)	**.26** (.25)	**.15** (.16)	**.14** (.11)	**.12** (.14)

Table C.3. (Continued)

WPS and WVI Scales	Rational Biodata Inventory Scales (Continued)						
	Cultural Tolerance	Internal Locus of Control	Army Identification	Respect Authority	Narcissism	Gratitude	Lie Scale
WVI Scales							
Social Status	.03 (.03)	.05 (.05)	**.14** (.14)	**.13** (.13)	**.10** (.10)	**.12** (.12)	-.01 (.01)
Advancement	**.12** (.12)	**.14** (.15)	.07 (.08)	**.13** (.13)	**.10** (.09)	**.14** (.15)	-.02 (-.03)
Autonomy	.03 (.04)	.03 (.07)	-.02 (-.01)	-.01 (-.01)	-.02 (-.03)	.01 (.04)	-.02 (-.05)
Supportive Supervision	.06 (.05)	.01 (-.01)	.05 (.05)	**.16** (.16)	.07 (.07)	**.10** (.09)	-.01 (.01)
Leisure Time	-.02 (-.01)	-.01 (.02)	**-.11** (-.10)	**-.09** (-.08)	-.04 (-.04)	.04 (.07)	-.07 (-.10)
Comfort	-.02 (-.02)	-.07 (-.06)	**-.20** (-.20)	-.03 (-.03)	-.01 (-.02)	.00 (.01)	-.07 (-.09)
Achievement	**.08** (.09)	.07 (.10)	.07 (.07)	**.11** (.12)	.05 (.04)	**.14** (.16)	-.07 (-.09)
Societal Contribution	**.11** (.12)	**.09** (.10)	**.10** (.10)	**.09** (.09)	-.02 (-.02)	**.16** (.17)	.01 (.00)
Independence	**-.14** (-.13)	-.07 (-.04)	**-.14** (-.13)	**-.15** (-.14)	-.03 (-.03)	**-.15** (-.13)	-.04 (-.06)
Social Service	**.14** (.14)	.06 (.06)	.08 (.07)	**.11** (.11)	-.05 (-.05)	**.19** (.18)	-.04 (-.04)
Fixed Role	.02 (.03)	.01 (.02)	.05 (. 06)	**.09** (.09)	.02 (.02)	.05 (.06)	-.04 (-.05)
Variety	**.10** (.10)	.05 (.06)	.00 (.01)	.04 (.04)	-.06 (-.06)	.04 (.06)	-.04 (-.05)
Leadership Opportunities	**.11** (.11)	**.10** (.10)	**.20** (.20)	**.15** (.15)	**.14** (.14)	**.12** (.12)	-.02 (-.02)
Feedback	.06 (.06)	.02 (.03)	.04 (.04)	**.11** (.12)	.03 (.03)	**.09** (.10)	-.05 (-.06)
Travel	**.10** (.10)	.03 (.03)	.04 (.04)	.01 (.01)	-.04 (-.04)	.01 (.01)	-.03 (-.03)
Physical Development	.06 (.06)	.06 (.05)	**.22** (.21)	**.11** (.11)	-.04 (-.03)	.13 (.13)	-.01 (-.01)
Ability Utilization	**.09** (.10)	**.09** (.13)	.07 (.08)	.03 (.04)	-.07 (-.07)	.09 (.12)	-.02 (-.06)
Creativity	.04 (.05)	-.01 (.02)	**-.13** (-.12)	-.07 (-.06)	.04 (.03)	-.01 (.02)	-.04 (.07)
Recognition	.01 (.01)	-.01 (.00)	.01 (.01)	.05 (.05)	**.10** (.10)	.00 (.01)	-.07 (-.07)
Co-Workers	**.12** (.12)	.06 (.08)	.01 (.01)	**.09** (.09)	-.03 (-.03)	**.19** (.21)	**-.08** (-.10)
Activity	.02 (.02)	.06 (.07)	.02 (.03)	.06 (.07)	**-.15** (-.15)	.01 (.02)	.03 (.01)
Flexible Schedule	-.01 (.00)	-.07 (-.05)	**-.16** (-.15)	**-.09** (-.08)	**-.12** (-.12)	-.02 (-.01)	**-.11** (-.12)
Personal Development	.07 (.07)	.06 (.07)	.03 (.04)	**.08** (.09)	-.05 (-.05)	**.09** (.10)	.01 (-.01)
Home	.00 (.01)	.03 (.06)	-.06 (-.05)	-.03 (-.02)	-.05 (-.06)	.04 (.07)	-.05 (-.08)
Esteem	.07 (.08)	.07 (.10)	.01 (.01)	.07 (.07)	.04 (.03)	.08 (.10)	-.02 (-.05)
Emotional Development	**.08** (.08)	**.10** (.12)	**.12** (.13)	.06 (.06)	-.02 (-.03)	.06 (.07)	-.04 (-.05)
Influence	-.02 (-.01)	.03 (.05)	.04 (.04)	.03 (.03)	.04 (.04)	-.01 (.01)	-.03 (-.04)
Team Orientation	**.17** (.17)	.09 (.10)	.02 (.03)	**.10** (.10)	-.07 (-.07)	**.15** (.15)	-.05 (-.05)

Note. Bold indicates $p < .05$, two-tailed for raw correlations. n = 487 – 672. Corrected correlations are in parentheses.

Table C.4. Correlations between WPS and WVI Scale Scores

	Work Preference Survey Scales										
WVI Scales	Realistic Interests	Mechanical Facet	Physical Facet	Investigative Interests	Critical Thinking Facet	Research Facet	Artistic Interests	Artistic Activities Facet	Creativity Facet	Social Interests	Work With Others Facet
Social Status	**.09** (.09)	.04 (.04)	**.10** (.10)	.01 (.01)	.04 (.04)	-.03 (-.03)	-.01(-.01)	-.01 (-.01)	.01 (.01)	**.16** (.16)	**.17** (.16)
Advancement	.05 (.04)	.05 (.04)	.05 (.04)	.05 (.06)	**.08** (.09)	.01 (.01)	.00 (.00)	.00 (-.01)	.00 (.01)	**.08** (.07)	.07 (.05)
Autonomy	.00 (-.04)	.03 (.01)	-.03 (-.06)	.02 (.05)	.06 (.10)	-.02 (-.02)	.02 (.01)	.00 (-.02)	.05 (.06)	**-.08** (-.11)	**-.09** (-.12)
Supportive Supervision	**.13** (.15)	**.13** (.14)	**.09** (.11)	.03 (.02)	.03 (.00)	.02 (.02)	.06 (.06)	**.07** (.08)	-.01 (-.01)	**.14** (.15)	**.14** (.15)
Leisure Time	.00 (-.03)	.05 (.03)	-.04 (-.07)	**-.08** (-.06)	-.06 (-.01)	**-.08** (-.08)	.01 (.00)	.02 (.00)	-.01 (.00)	**-.11** (-.13)	**-.08** (-.11)
Comfort	-.05 (-.07)	.06 (.05)	**-.15** (-.16)	-.06 (-.05)	**-.09** (-.07)	-.02 (-.02)	**.08** (.08)	**.11** (.10)	-.01 (-.01)	-.05 (-.05)	**-.08** (-.09)
Achievement	.02 (-.01)	.05 (.04)	-.01 (-.03)	**.08** (.10)	**.09** (.12)	.04 (.04)	.06 (.05)	.04 (.02)	**.09** (.10)	**.09** (.07)	.04 (.01)
Societal Contribution	.02 (.01)	.00 (-.01)	.05 (.04)	**.09** (.10)	**.11** (.12)	.06 (.06)	.06 (.05)	.05 (.05)	.04 (.04)	**.24** (.23)	**.14** (.13)
Independence	-.01 (-.03)	**.09** (.07)	**-.11** (-.13)	-.06 (-.04)	-.06 (-.02)	-.04 (-.04)	.06 (.05)	.07 (.05)	.02 (.03)	**-.26** (-.27)	**-.27** (-.29)
Social Service	.04 (.04)	.03 (.03)	.03 (.04)	**.10** (.09)	**.12** (.11)	.05 (.05)	.04 (.04)	.03 (.03)	.05 (.05)	**.32** (.32)	**.18** (.18)
Fixed Role	.07 (.06)	**.10** (.09)	.03 (.02)	.02 (.02)	.07 (.08)	-.04 (-.04)	-.05 (-.05)	-.03 (-.04)	-.07 (-.06)	01 (.00)	.01 (.00)
Variety	**.19** (.18)	**.19** (.18)	**.12** (.11)	.01 (.02)	.05 (.07)	-.04 (-.04)	.06 (.05)	.03 (.03)	**.09** (.09)	.04 (.03)	.07 (.05)
Leadership Opportunities	**.11** (.11)	.07 (.07)	**.13** (.13)	**.14** (.14)	**.21** (.20)	.04 (.04)	.05 (.05)	.02 (.02)	**.10** (.10)	**.24** (.24)	**.22** (.22)
Feedback	.05 (.04)	**.10** (.09)	-.02 (-.02)	.03 (.03)	**.07** (.08)	-.02 (-.02)	.04 (.04)	.03 (.03)	.05 (.05)	.06 (.06)	.04 (.04)
Travel	**.11** (.11)	.07 (.07)	**.11** (.10)	.06 (.05)	**.08** (.07)	.02 (.02)	**.10** (.10)	**.10** (.10)	.07 (.07)	.02 (.02)	.07 (.07)
Physical Development	**.30** (.30)	**.14** (.14)	**.39** (.40)	-.03 (-.03)	.01 (.00)	-.05 (-.05)	-.01 (.00)	.02 (.02)	-.06 (-.06)	.06 (.06)	**.12** (.12)
Ability Utilization	**.07** (.03)	**.13** (.10)	.00 (-.04)	**.11** (.14)	**.16** (.20)	.04 (.04)	**.08** (.07)	.05 (.03)	**.10** (.11)	.03 (.00)	.00 (-.04)
Creativity	-.01 (-.05)	**.09** (.06)	**-.10** (-.13)	**.09** (.11)	**.10** (.15)	.05 (.05)	**.21** (.20)	**.16** (.14)	**.22** (.23)	-.01 (-.04)	-.05 (-.08)
Recognition	.01 (.00)	.04 (.03)	-.04 (-.05)	-.03 (-.02)	.01 (.02)	-.05 (-.05)	.02 (.02)	.02 (.01)	.01 (.02)	.02 (.01)	.02 (.01)
Co-Workers	.06 (.04)	**.08** (.06)	.02 (.00)	-.01 (.01)	.02 (.05)	-.04 (-.03)	.06 (.05)	.06 (.05)	.03 (.03)	**.12** (.10)	**.15** (.12)
Activity	**.10** (.08)	**.15** (.14)	.00 (-.01)	.00 (.01)	.05 (.06)	-.05 (-.05)	-.03 (-.03)	-.03 (-.04)	-.01 (-.01)	-.04 (-.05)	-.04 (-.05)
Flexible Schedule	-.03 (-.05)	.05 (.03)	**-.09** (-.11)	**-.13** (-.11)	**-.12** (-.10)	**-.09** (-.09)	.02 (.01)	.05 (.04)	-.05 (-.04)	**-.13** (-.14)	**-.11** (-.13)
Personal Development	**.08** (.07)	**.13** (.12)	.01 (.00)	.07 (.08)	**.12** (.13)	.00 (.00)	.00 (.00)	-.01 (-.02)	.03 (.03)	.04 (.03)	.05 (.04)
Home	-.05 (-.08)	-.01 (-.03)	-.06 (-.09)	-.04 (-.02)	-.02 (.03)	-.05 (-.05)	-.03 (-.04)	-.02 (-.03)	-.05 (-.03)	-.02 (-.04)	-.05 (-.08)
Esteem	.00 (-.03)	.03 (.01)	-.03 (-.06)	.06 (.08)	**.11** (.15)	.00 (.00)	.01 (.00)	.01 (-.01)	.02 (.03)	.03 (.00)	.03 (.00)
Emotional Development	**.14** (.13)	**.10** (.09)	**.15** (.14)	.05 (.06)	**.11** (.13)	-.03 (-.02)	-.05 (-.06)	-.05 (-.05)	-.04 (-.04)	.07 (.06)	**.09** (.08)
Influence	.03 (.01)	.05 (.03)	.01 (-.01)	.07 (.08)	**.10** (.12)	.01 (.02)	-.01 (-.01)	-.02 (-.03)	.02 (.03)	.03 (.01)	.03 (.01)
Team Orientation	.03 (.02)	.01 (.00)	.04 (.03)	.02 (.03)	.03 (.04)	.01 (.01)	.05 (.05)	.06 (.06)	.01 (.02)	**.18** (.17)	**.19** (.18)

Table C.4. (Continued)

WVI Scales	Work Preference Survey Scales Continued								
	Help Others Facet	Enterprising Interests	Prestige Facet	Lead Others Facet	High Profile Facet	Conventional Interests	Information Management Facet	Detail Orientation Facet	Clear Procedures Facet
Social Status	**.10** (.10)	**.16** (.16)	**.21** (.21)	**.17** (.17)	.01 (.01)	**.11** (.11)	.04 (.04)	**.09** (.09)	**.12** (.12)
Advancement	.03 (.02)	**.16** (.15)	**.20** (.20)	**13** (.12)	.06 (.05)	**.14** (.13)	**.08** (.08)	**.11** (.10)	**.15** (.14)
Autonomy	-.06 (-.07)	.00 (-.02)	.02 (.02)	-.01 (-.04)	-.02 (-.04)	-.04 (-.07)	-.04 (-.06)	.02 (-.01)	.02 (-.02)
Supportive Supervision	**.08** (.08)	**.09** (.09)	**.10** (.10)	**.08** (.09)	.05 (.06)	**.20** (.21)	**.12** (.13)	.11 (.12)	**.16** (.18)
Leisure Time	**-.11** (-.12)	**-.10** (-.11)	-.02 (-.03)	**-.14** (-.16)	-.06 (-.08)	**-.12** (-.14)	**-.10** (-.11)	**-.10** (-.12)	-.07 (-.10)
Comfort	-.01 (-.01)	-.06 (-.07)	.00 (.00)	**-.13** (-.14)	.00 (.00)	.02 (.01)	.05 (.05)	-.06 (-.07)	-.02 (-.03)
Achievement	**.09** (.08)	.06 (.05)	**.12** (.12)	.01 (-.02)	-.01 (-.02)	.06 (.03)	.00 (-.01)	**.10** (.08)	**.10** (.07)
Societal Contribution	**.24** (.24)	.07 (.06)	.04 (.03)	**.08** (.07)	.02 (.02)	.05 (.04)	-.02 (-.03)	**.10** (.09)	**.11** (.09)
Independence	**-.18** (-.19)	**-.10** (-.11)	**-.09** (-.09)	**-.19** (-.21)	.00 (-.02)	-.07 (-.10)	-.06 (-.07)	-.05 (-.07)	-.04 (-.07)
Social Service	**.35** (.35)	.07 (-.07)	.05 (.05)	**13** (.14)	-.01 (.00)	**.11** (.12)	.03 (.03)	**.12** (.12)	**.16** (.16)
Fixed Role	-.01 (-.01)	.03 (.03)	.07 (.07)	.04 (.03)	-.01 (-.02)	**.19** (.18)	.06 (.06)	**13** (.12)	**.21** (.20)
Variety	-.03 (-.03)	-.03 (-.03)	.01 (.01)	.01 (-.01)	-.07 (-.07)	-.05 (-.06)	**-.09** (-.10)	.03 (.02)	.03 (.01)
Leadership Opportunities	**.15** (.15)	**.29** (.29)	**.23** (.23)	**37** (.36)	**.10** (.10)	**.15** (.15)	.06 (.06)	**.16** (.16)	**.18** (.18)
Feedback	.04 (.04)	.04 (.03)	**.08** (.08)	.02 (.01)	.00 (.00)	**.09** (.09)	.04 (.03)	**.09** (.09)	**.11** (.11)
Travel	-.05 (-.05)	.05 (.05)	.01 (.01)	.01 (.01)	.06 (.06)	-.04 (-.04)	-.07 (-.07)	.06 (.06)	.02 (.02)
Physical Development	-.04 (-.03)	.01 (.02)	.02 (.02)	.06 (.07)	-.03 (-.02)	**.10** (.10)	.02 (.02)	**.10** (.11)	**13** (.13)
Ability Utilization	.02 (.00)	-.01 (-.03)	.03 (.02)	-.03 (-.07)	-.05 (-.07)	.07 (.03)	.02 (-.01)	**.14** (.11)	**.10** (.06)
Creativity	-.01 (-.02)	.03 (.01)	.04 (.03)	-.06 (-.09)	.04 (.02)	-.01 (-.04)	.01 (-.01)	.01 (-.01)	-.02 (-.05)
Recognition	-.01 (-.01)	**.10** (.10)	**.21** (.21)	.02 (.01)	.02 (.02)	.06 (.06)	.03 (.02)	.03 (.02)	.06 (.05)
Co-Workers	.06 (.05)	-.01 (-.02)	.02 (.01)	.00 (-.02)	-.02 (-.03)	.03 (.01)	.02 (.01)	-.01 (-.03)	.01 (-.01)
Activity	-.04 (-.05)	-.07 (-.07)	-.01 (-.02)	-.04 (-.05)	**-.09** (-.09)	.07 (.06)	-.02 (-.02)	**.15** (.14)	13 (.12)
Flexible Schedule	**-.09** (-.10)	**-.13** (-.14)	-.07 (-.07)	**-.18** (-.20)	-.04 (-.05)	**-.09** (-.10)	-.05 (-.06)	**-.10** (-.12)	-.07 (-.08)
Personal Development	-.01 (-.01)	-.01 (-.02)	.03 (.02)	-.01 (-.02)	-.04 (-.05)	**.11** (.09)	.04 (.03)	**.14** (.13)	**.14** (.13)
Home	.02 (.01)	-.05 (-.06)	.01 (.01)	**-.08** (-.11)	-.04 (-.05)	-.01 (-.04)	-.03 (-.04)	-.02 (-.04)	.03 (-.01)
Esteem	.01 (.00)	.07 (.05)	**.12** (.12)	.01 (-.02)	.03 (.02)	**.08** (.05)	.05 (.03)	.07 (.05)	**.09** (.06)
Emotional Development	.00 (.00)	.03 (.02)	.03 (.03)	**.09** (.07)	-.03 (-.04)	**.12** (.10)	.02 (.02)	**.16** (.14)	**.16** (.14)
Influence	.00 (-.01)	**.11** (.10)	.07 (.06)	**.11** (.09)	.05 (.04)	**.11** (.09)	.05 (.04)	**.12** (.11)	**.12** (.10)
Team Orientation	**.12** (.12)	.02 (.02)	.00 (.00)	.04 (.03)	.01 (.01)	.05 (.04)	.05 (.04)	.02 (.02)	.02 (.01)

Note. Bold indicates $p < .05$, two-tailed for raw correlations. n = 707 - 766. Corrected correlations are in parentheses.

Table C.5. Composite Intercorrelations

		WSI										WPS		WVI	
Instrument/Composite		1	2	3	4	5	6	7	8	9	10	11	12	13	14
WSI															
1	Predictor for Expected Future Performance	--	.50	.46	.12	.03	.18	.20	-.24	.13	.25	.09	.01	.10	.08
2	Predictor for General Technical Proficiency	**.50**	--	.26	.04	.19	.07	.24	-.13	.10	.09	.08	-.06	-.04	-.06
3	Predictor for Achievement and Effort	**.46**	**.27**	--	.22	-.07	.34	.33	-.39	.38	.35	.08	.05	.20	.19
4	Predictor for Physical Fitness	**.12**	.05	**.22**	--	-.39	.59	.62	-.42	.57	.41	.05	.21	.18	.27
5	Predictor for Teamwork	.03	**.19**	-.07	**-.39**	--	-.46	-.42	.55	-.34	-.52	-.02	-.17	-.13	-.23
6	Predictor for Satisfaction with the Army	**.19**	**.10**	**.34**	**.59**	**-.46**	--	.86	-.63	.68	.57	.14	.26	.29	.38
7	Predictor for Perceived Army Fit	**.21**	**.26**	**.33**	**.62**	**-.42**	**.86**	--	-.60	.75	.62	.16	.27	.25	.36
8	Predictor for Attrition Cognitions	**-.24**	**-.12**	**-.39**	**-.42**	**.55**	**-.64**	**-.61**	--	-.54	-.50	-.10	-.15	-.19	-.28
9	Predictor for Career Intentions	**.14**	**.12**	**.38**	**.57**	**-.34**	**.68**	**.75**	**-.55**	--	.48	.13	.24	.25	.35
10	Predictor for Future Army Affect	**.25**	**.09**	**.35**	**.41**	**-.52**	**.57**	**.62**	**-.50**	**.48**	--	.11	.18	.16	.27
WPS															
11	Unit Achievement and Effort	**.08**	**.07**	**.08**	**.05**	-.02	**.15**	**.17**	**-.10**	**.14**	**.11**	--	.62	.33	.29
12	Subjective Perceived Army Fit	.03	-.01	.05	**.21**	**-.17**	**.25**	**.26**	**-.16**	**.23**	**.18**	**.65**	--	.42	.47
WVI															
13	Unit Achievement and Effort	**.11**	-.02	**.20**	**.18**	**-.13**	**.28**	**.24**	**-.19**	**.25**	**.16**	**.34**	**.41**	--	.73
14	Unit Satisfaction with the Army	**.09**	-.03	**.19**	**.27**	**-.23**	**.37**	**.36**	**-.29**	**.35**	**.27**	**.30**	**.46**	**.73**	--

Note. Bold indicates $p < .05$ for raw correlations. n = 640 - 732. Raw correlations appear below the diagonal. Corrected correlations are above the diagonal.

Table C.6. Correlations between WSI Composites and ASVAB, Target Tracking, PSJT, and RBI Scale Scores

	Work Suitability Inventory Composites									
	FXP	GTP	AE	PF	TEAM	ASat	AFit	ACog	CInt	FAA
ASVAB										
AFQT	.05 (.08)	**.20** (.31)	-.01(-.01)	-.02 (-.03)	.02 (.03)	**-.08** (-.13)	-.06 (-.09)	-.04 (-.06)	-.07 (-.11)	-.01 (-.02)
Spatial	.08 (.10)	.04 (.13)	-.01 (-.01)	.03 (.02)	.00 (.01)	-.01 (-.05)	-.02 (-.05)	-.01 (-.03)	-.06 (-.09)	-.03 (-.03)
Technical	.07 (.09)	**.20** (.29)	-.03 (-.03)	.03 (.02)	.00 (.02)	-.04 (-.09)	.02 (-.02)	-.11 (-.12)	-.09 (-.12)	.00 (-.01)
Target Tracking	.04 (.06)	**.10** (.15)	.03 (.02)	.04 (.04)	.01 (.01)	.03 (.01)	.04 (.02)	-.04 (-.05)	-.02 (-.04)	.04 (.04)
PSJT Judgment	.04 (.06)	.02 (.08)	**.10** (.10)	.03 (.02)	.05 (.05)	.06 (.04)	.03 (.01)	.02 (.01)	.02 (.00)	-.01 (-.01)
RBI (*lie adjusted*)										
Peer Leadership	**.14** (.14)	**.21** (.24)	.05 (.05)	**.12** (.11)	.02 (.03)	**.10** (.08)	.15 (.14)	-.09 (-.10)	.11 (.09)	.11 (.11)
Cognitive Flexibility	**.08** (.10)	**.16** (.23)	.06 (.05)	.00 (-.01)	.03 (.04)	-.06 (-.09)	-.01 (-.04)	-.01 (-.02)	-.01 (-.04)	.07 (.06)
Achievement	**.15** (.15)	**.10** (.12)	**.19** (.19)	**.19** (.19)	-.03 (-.03)	**.22** (.21)	.24 (.23)	-.17 (-.17)	.24 (.23)	.17 (.16)
Fitness Motivation	.04 (.05)	.10 (.10)	**.16** (.15)	**.26** (.26)	**-.20** (-.20)	**.29** (.28)	**.35** (.35)	**-.22** (-.22)	**.27** (.26)	**.30** (.30)
Interpersonal Skills – Diplomacy	-.04 (-.04)	-.03 (-.02)	.02 (.02)	**.14** (.14)	**-.10** (-.10)	**.14** (.14)	.10 (.09)	.00 (-.01)	.05 (.05)	.06 (.06)
Stress Tolerance	-.01 (.00)	.06 (.10)	.02 (.02)	.02 (.01)	-.08 (-.07)	**.10** (.08)	.09 (.08)	-.02 (-.03)	.03 (.01)	.12 (.12)
Hostility to Authority	.05 (.04)	.05 (.00)	**-.11** (-.11)	.02 (.02)	-.04 (-.05)	-.01 (.01)	.01 (.02)	-.04 (-.04)	-.01 (.01)	.02 (.02)
Self-Efficacy	**.15** (.16)	**.17** (.21)	**.13** (.13)	**.14** (.14)	-.08 (-.07)	**.18** (.16)	.21 (.20)	-.18 (-.18)	.13 (.11)	.20 (.19)
Cultural Tolerance	-.05 (-.04)	-.03 (-.01)	.02 (.02)	.02 (.02)	-.01 (-.01)	**.09** (.08)	.07 (.07)	.09 (.09)	.06 (.05)	.13 (.12)
Internal Locus of Control	.05 (.06)	.01 (.05)	**.18** (.17)	.05 (.04)	-.07 (-.06)	**.15** (.13)	.12 (.10)	-.12 (-.13)	.11 (.09)	.13 (.13)
Army Identification	**.10** (.10)	**.11** (.11)	**.16** (.16)	**.27** (.27)	**-.21** (-.21)	**.33** (.33)	.38 (.37)	-.27 (-.27)	.31 (.30)	.28 (.28)
Respect for Authority	.06 (.06)	-.06 (-.06)	**.08** (.08)	**.16** (.16)	**-.10** (-.10)	**.19** (.19)	.19 (.19)	-.16 (-.16)	.19 (.19)	.11 (.11)
Narcissism	.06 (.06)	.04 (.03)	.07 (.07)	**.09** (.09)	-.05 (-.05)	**.08** (.08)	.10 (.10)	-.09 (-.09)	.11 (.11)	.12 (.12)
Gratitude	.01 (.02)	-.04 (-.01)	.02 (.02)	**.12** (.11)	**-.09** (-.08)	**.17** (.16)	.10 (.09)	-.08 (-.09)	.11 (.09)	.06 (.05)
Lie Scale	-.04 (-.05)	-.03 (-.07)	.01 (.01)	.04 (.05)	-.01 (-.01)	.05 (.07)	.06 (.07)	-.04 (-.03)	.04 (.05)	.03 (.03)

Note. Bold indicates $p < .05$. $n = 487 – 653$. EXP = WSI Empirical Dyad Composite (EDC) for Future Expected Performance (FXP), GTP = EDC General Technical Proficiency, AE = EDC Achievement and Effort, PF = EDC Physical Fitness, TEAM = EDC Teamwork, ASat = EDC Satisfaction with the Army, AFit = EDC Perceived Army Fit, ACog = EDC Attrition Cognitions, CInt = EDC Career Intentions, FAA = EDC Future Army Affect.

Table C.7. Correlations between WPS and WVI Composites with ASVAB, Target Tracking, PSJT, and RBI Scale Scores

	WPS Composite		WVI Composite	
	Unit Achievement and Effort	Subjective Perceived Army Fit	Unit Achievement and Effort	Unit Satisfaction with the Army
ASVAB				
AFQT	.04 (.06)	**-.16** (-.25)	**-.08** (-.13)	**-.11** (-.18)
Spatial	.00 (.02)	-.03 (-.11)	-.05 (-.09)	**-.10** (-.14)
Technical	.01 (.04)	**-.09** (-.18)	**-.14** (-.17)	**-.13** (-.18)
Target Tracking	.02 (.03)	.01 (-.03)	.01 (-.02)	.02 (-.01)
PSJT Judgment	**.29** (.30)	**.18** (.12)	**.25** (.21)	**.16** (.11)
RBI (*lie adjusted*)				
Peer Leadership	**.32** (.33)	**.27** (.22)	**.15** (.12)	**.13** (.10)
Cognitive Flexibility	**.40** (.39)	**.29** (.19)	**.17** (.12)	**.13** (.07)
Achievement	**.45** (.45)	**.43** (.41)	**.36** (.35)	**.33** (.32)
Fitness Motivation	**.22** (.22)	**.36** (.35)	**.16** (.16)	**.28** (.27)
Interpersonal Skills - Diplomacy	**.29** (.29)	**.33** (.32)	**.21** (.20)	**.24** (.23)
Stress Tolerance	.04 (.05)	.02 (-.01)	.06 (.04)	**.11** (.08)
Hostility to Authority	**-.20** (-.20)	-.06 (-.02)	**-.15** (-.13)	**-.13** (-.10)
Self-Efficacy	**.36** (.36)	**.32** (.27)	**.18** (.16)	**.23** (.20)
Cultural Tolerance	**.28** (.29)	**.25** (.23)	**.21** (.20)	**.19** (.18)
Internal Locus of Control	**.26** (.26)	**.20** (.16)	**.17** (.15)	**.21** (.18)
Army Identification	**.22** (.22)	**.38** (.36)	**.30** (.30)	**.42** (.41)
Respect for Authority	**.30** (.30)	**.35** (.34)	**.23** (.23)	**.27** (.26)
Narcissism	**.13** (.13)	**.19** (.19)	.07 (.07)	.06 (.06)
Gratitude	**.20** (.21)	**.28** (.24)	**.21** (.19)	**.21** (.19)
Lie Scale	**.08** (.07)	**.08** (.10)	.03 (.04)	.06 (.08)

Note. Bold indicates $p < .05$. $n = 553 – 738$. Corrected correlations are in parentheses.